MCB
Journals.ASM.org

Novel Long Noncoding RNAs (lncRNAs) in Myogenesis: a *miR-31* Overlapping lncRNA Transcript Controls Myoblast Differentiation

Monica Ballarino,[a] Valentina Cazzella,[a] Daniel D'Andrea,[b] Luigi Grassi,[b] Lavinia Bisceglie,[a] Andrea Cipriano,[a] Tiziana Santini,[c] Chiara Pinnarò,[a] Mariangela Morlando,[a] Anna Tramontano,[b,c] Irene Bozzoni[a,c,d]

Dept. of Biology and Biotechnology Charles Darwin, Sapienza University of Rome, Rome, Italy[a]; Dept. of Physics, Sapienza University of Rome, Rome, Italy[b]; Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Rome, Italy[c]; Institute Pasteur Fondazione Cenci-Bolognetti and IBPM, Sapienza University of Rome, Rome, Italy[d]

Transcriptome analysis allowed the identification of new long noncoding RNAs differentially expressed during murine myoblast differentiation. These transcripts were classified on the basis of their expression under proliferating versus differentiated conditions, muscle-restricted activation, and subcellular localization. Several species displayed preferential expression in dystrophic (*mdx*) versus wild-type muscles, indicating their possible link with regenerative processes. One of the identified transcripts, *lnc-31*, even if originating from the same nuclear precursor of *miR-31*, is produced by a pathway mutually exclusive. We show that *lnc-31* and its human homologue *hsa-lnc-31* are expressed in proliferating myoblasts, where they counteract differentiation. In line with this, both species are more abundant in *mdx* muscles and in human Duchenne muscular dystrophy (DMD) myoblasts, than in their normal counterparts. Altogether, these data suggest a crucial role for *lnc-31* in controlling the differentiation commitment of precursor myoblasts and indicate that its function is maintained in evolution despite the poor sequence conservation with the human counterpart.

Earlier investigations of the muscle development of vertebrates led to the discovery of transcription factors (MYOD1, MYF5, and MYOG) and chromatin modifiers as key regulators of muscle specification (1–4). Advancement in the field was stimulated by the identification of regulatory small noncoding RNAs (sncRNAs) and long noncoding RNAs (lncRNAs) as critical orchestrators of muscle specification (5, 6). Among sncRNAs, the characterization of *miR-1*/*miR-206* and *miR-133* microRNA (miRNA) families in both muscle physiology and diseases was of primary interest (7–10). Besides characterizing miRNAs, a number of studies also focused on the participation of lncRNAs in muscular circuitries (11–13). lncRNAs represent a class of transcripts longer than 200 nucleotides with little or no protein-coding capacity. They regulate gene expression through the formation of different ribonucleoprotein complexes and exert functions in a variety of cellular pathways (5, 14). *linc-MD1* was the first example of a cytoplasmic lncRNA guiding the timing of mouse muscle differentiation *in vitro* (11). During differentiation of myoblasts, *linc-MD1* acts as a sponge for *miR-133* and *miR-135*, thereby derepressing their myogenic targets, *Maml1* and *Mef2c*. Notably, the levels of human *linc-MD1* expression are low in myoblasts of patients affected by Duchenne muscular dystrophy (DMD) (15), while proper differentiation was rescued by exogenous administration of *linc-MD1* (11). More recently, other muscle-specific lncRNAs, containing short interspersed elements (SINEs) and regulating gene expression by driving Staufen1 (STAU1)-mediated mRNA decay, were also linked to myogenesis (13). However, the number of lncRNAs identified as myogenic regulators so far is still exiguous and the identification of new species is required to define the role of these transcripts in muscle specification.

In this study, we characterized novel lncRNA species, including previously unannotated transcripts, which undergo modulated expression during *in vitro* murine myoblast differentiation. Parallel tissue-specific analysis identified a subgroup of lncRNAs preferentially expressed in mature or in regenerating fibers. Specific binding of key myogenic transcription factors on the putative promoter regions of those species upregulated during differentiation was then identified.

The resulting characterization of *lnc-31* was particularly interesting due to its genomic overlapping with the *miR-31* coding region. Previously linked to neoplastic development and tumor metastasis (16), *miR-31* was shown to maintain satellite cell replication through the posttranscriptional control of MYF5 (17) and to repress late myogenic markers, including dystrophin (7). Here we explain that, in muscle cells, *miR-31* originates from a primary transcript specifically expressed under proliferating conditions that can be converted into *miR-31* or into the mature polyadenylated *lnc-31* species. By modulating the levels of *lnc-31*, we demonstrated a relevant role for this molecule in controlling the maintenance of myoblast proliferation both in murine and human myoblasts.

MATERIALS AND METHODS

Cell culture conditions and treatments. C_2C_{12} cells were cultured as previously described (7). For differentiation, cells maintained under confluent conditions were switched to a low serum concentration (0.5% fetal bovine serum [FBS]). Human myoblasts (WT-9808 and DMD-9981 from

Received 17 November 2014 **Accepted** 5 December 2014

Accepted manuscript posted online 15 December 2014

Citation Ballarino M, Cazzella V, D'Andrea D, Grassi L, Bisceglie L, Cipriano A, Santini T, 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. Mol Cell Biol 35:728–736. doi:10.1128/MCB.01394-14.

Address correspondence to Irene Bozzoni, irene.bozzoni@uniroma1.it.

M.B. and V.C. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/MCB.01394-14.

doi:10.1128/MCB.01394-14

Telethon Neuromuscular Biobank) were cultured and induced to differentiate as previously described (18). Small interfering RNA (siRNA) molecules (Qiagen) against *lnc-31* exon2 (DNA target sequence: 5′-TCACGT TGTTGAAGAGTTGAA-3′) were transfected using HiPerfect (Qiagen) according to manufacturer's specifications. siRNA molecules against human *lnc-31* (a gift from Anders Lund) were transfected into human proliferating myoblasts using Lipofectamine 2000 (Life Technologies) according to the manufacturer's specifications.

mRNA library generation and RNA-seq. Transcriptome sequencing (RNA-seq) libraries were prepared from 200 ng of total RNA using a TruSeq stranded mRNA Sample Prep kit (Illumina) and following manufacturer's instructions. DNA libraries were quantified with Kapa reagent (Kapa SYBR Fast universal quantitative PCR [qPCR] kit; Illumina) and inspected for quality using a 2100 Bioanalyzer (Agilent). Multiplexed libraries (10 pM) were sequenced on a Genome Analyzer IIx (GAIIx) system (Illumina Inc.) as paired-end (2 × 86) base reads at about 50 million mapped reads per sample.

RNA-seq data analysis. Paired-end (86-bp) reads were aligned to the mouse reference genome assembly (mm9) using Tophat (19) with default options and assembled into transcripts with Cufflinks (20, 21). Cuffmerge with default options was used to merge all the assembled transcriptomes. The aligned reads and the assembled transcriptomes were used as input for Cuffdiff2 (22) to determine the expression levels in fragments per kilobase per million (FPKM). FPKM values of the newly identified lncRNAs are reported in Table S1 in the supplemental material. Cells at all the differentiation stages (differentiating myoblasts on day 1 [DM1], DM3, and DM5) were compared with undifferentiated cells (i.e., growing myoblasts [GM]). We considered all transcripts with a FPKM value of >0.1 to represent expressed transcripts (23, 24). In each pairwise comparison, we calculated the fold change for transcripts expressed in each of the two samples and selected those with a fold change value greater than 1. The resulting list was used to compute a threshold value corresponding to the third quartile of the data. Transcripts with fold change values above the threshold were ordered according to the corresponding fold change value. Transcripts not expressed (or with an FPKM value lower than 0.1) in one of the samples and with a FPKM value greater than 1 in the other were ordered according to the FPKM value for the latter sample. The same procedure was used to select the genes for functional classification.

Animal procedures. The animals were treated according to the guidelines of good laboratory practice (GLP) with respect to housing, nutrition, and care. Two-month-old wild-type (WT) and dystrophic (*mdx*) animals (C57black6) were sacrificed, and tissues from different body districts were homogenized into powder under liquid nitrogen by mortar and pestle.

RNA analyses. Total RNA was prepared with an RNeasy Plus minikit (Qiagen); 0.5 to 1.0 μg of total RNA was subjected to reverse transcription (RT) with either SuperScript III reverse transcriptase (Life Technologies) or miScript reverse transcription (Qiagen) according to the manufacturer's instructions. Semiquantitative PCR was then performed using MyTaq (Bioline) enzyme. All quantitative RT-PCRs (qRT-PCRs) were performed in triplicate using a SYBR green PCR kit (Qiagen). Cytoplasmic, nucleoplasmic, and chromatin fractionation was carried out as described in reference 25, and lncRNAs were analyzed in GM or DM (day 2), depending on their timing of expression. Northern blot analysis for *miR-31* was performed as described in reference 7 using a *miR-31* locked nucleic acid (LNA) probe. For primer specifications, refer to Table S1 in the supplemental material.

Western blot analyses. Total protein extract was obtained as previously described (26). Immunoblots were incubated with the following primary antibodies: anti-MYH6 (anti-MF-20), anti-muscle creatine kinase (sc-15161; Santa Cruz Biotechnology, Inc., Santa Cruz, CA), antimyogenin (sc-12732; Santa Cruz), anti-MYOD1 (Dako, Glostrup, Denmark), and anti-GAPDH (anti-glyceraldehyde-3-phosphate dehydrogenase) (6C5 [sc-32233]; Santa Cruz). ImmunoPureGoat anti-rabbit or anti-mouse IgG–peroxidase (Pierce, Rockford, IL) (conjugated) was used as a secondary antibody.

BrdU assay. Proliferating C_2C_{12} and human myoblasts were labeled with bromodeoxyuridine (BrdU) labeling reagent (5-bromo-2′-deoxyuridine labeling and detection kit; Roche) for 3 h. Cells were fixed in an ice-cold solution (7 volumes of 100% ethanol and 3 volumes of 50 mM glycine, pH 2) for 20 min at −20°C and washed three times with phosphate-buffered saline (PBS). Fixed cells were incubated for 1 h at 37°C with anti-BrdU antibody. Cells were stained with a secondary antibody and then labeled with DAPI (4′,6-diamidino-2-phenylindole). Samples were imaged using an Axio Observer A1 (Zeiss) microscope.

Accession number. Deep-sequencing data have been submitted to the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena/data/view/PRJEB6112) under accession number PRJEB6112.

RESULTS AND DISCUSSION

Profiling and functional classification of differentially expressed genes. Transcriptome analysis was performed with RNA from murine C_2C_{12} myoblasts induced to differentiate *in vitro*. Proper differentiation timing was monitored by the microscopic observation of fiber formation (see Fig. S1A in the supplemental material) and by observation of the expression of known myogenic markers (27, 28) as MYOD1, MYOG, MEF2C, MYH6, and CKM (Fig. 1A).

Poly(A)-positive [poly(A)$^+$] RNA from proliferating (GM) and differentiating (DM1, DM3, and DM5) myoblasts was subjected to RNA-seq analysis. Confidence in the sequencing experiment was instilled by corresponding FPKM values of *Myod1*, *Myog*, *Mef2c*, *Myh6*, and *Ckm* transcripts (see Fig. S1B in the supplemental material) and by the appearance of their respective proteins throughout the time course of C_2C_{12} differentiation. (Fig. 1A).

The reconstructed transcriptome consisted of 72,326 expressed RNAs corresponding to 22,115 unique gene loci. About 68% of them were annotated in the Ensembl reference transcriptome (Fig. 1C), 23% corresponded to new isoforms of known transcripts, and 1% matched with antisense transcripts or exhibited a partial overlap to known genes. The remaining 7% were novel species not corresponding to any of the Ensembl reference transcripts. The gene ontology (GO) analysis was performed on a subset of protein-coding genes (see Fig. S1D in the supplemental material) using FIDEA (29). In line with the expected acquisition of a muscle signature, the GO diagram corresponding to DM conditions reveals an overrepresentation of myogenic transcripts, among them those controlling muscle system process, muscle contraction, and muscle structure development.

RNA sequencing data were then specifically inspected for lncRNAs differentially expressed during C_2C_{12} muscle differentiation. We focused on multiexonic transcripts more than 200 nucleotides long and discarded those overlapping known mRNA exons in the same strand. Among the top ranked, a subset of 30 lncRNAs (see Table S1 in the supplemental material) was selected on the basis of their expression levels, with an FPKM value of >1 in at least one of the differentiation time points. Of these species, 70% had FPKM values of >5, with some reaching FPKM values between 20 and 50; 18 lncRNAs were identified as new species, not previously annotated in the available databases. Therefore, we obtained a list of selected lncRNAs which were produced at meaningful levels and modulated during differentiation.

Sequencing data were confirmed by semiquantitative RT-PCR (sqRT-PCR) followed by sequencing of the amplified products: all of the selected lncRNAs were produced from the correct genomic locus. Twenty-three of such candidates were upregulated during

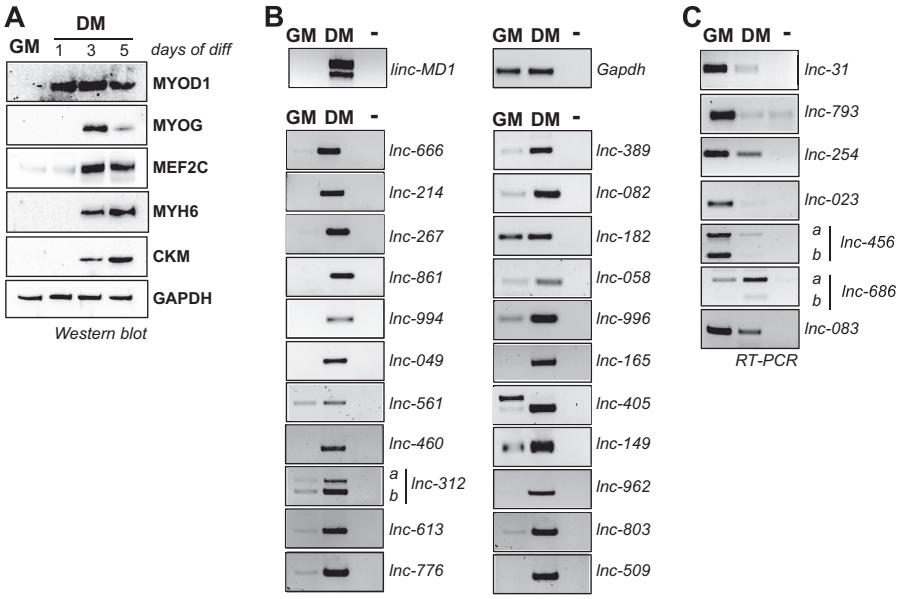

FIG 1 Muscle-specific poly(A)$^+$ RNA-seq profiling. (A) Western blot analysis of MYOD1, MYOG, MEF2C, MYH6, and CKM protein expression in myoblasts grown under proliferation (GM) and differentiation (DM; 1, 3, and 5 days of differentiation [diff]) conditions. GAPDH levels were used as a control. (B and C) Validation of lncRNA expression by semiquantitative RT-PCR (sqRT-PCR) performed on total RNA extracted from GM or from myoblasts differentiated for 3 days (DM). lncRNA transcripts were grouped into upregulated (B) and downregulated (C) categories. −, RT minus control reactions. *Gapdh* mRNA was used as control. "a" and "b" indicate *lnc-312*, *lnc-456*, and *lnc-686* splicing isoforms. PCR amplifications were performed on biological replicates, and the results of a representative experiment are shown.

the switch from GM to DM (Fig. 1B), while seven were specifically downregulated during differentiation (Fig. 1C). Notably, among the upregulated species, we found *linc-MD1* (Fig. 1B, upper panel), previously shown to be an abundant species that plays a relevant role in muscle differentiation. (11, 30).

To further analyze the noncoding features of these transcripts, we used a combination of three *in silico* approaches, Coding Potential Calculator (CPC) (31), PORTRAIT (32), and CPAT (33) (see Fig. S1E in the supplemental material). Our pipeline was first validated by inclusion of noncoding (*linc-MD1*, *Hotair*, *Malat1*, *H19*, and *Xist*) and coding (myoglobin, alpha hemoglobin, *Gapdh*, and *Acta2*) transcripts (see Fig. S1E and Table S2). When these criteria were applied to the newly identified transcripts (represented by gray dots in Fig. S1E), all of them lacked coding potential with the exception of four lncRNAs (*lnc-058*, *lnc-165*, *lnc-214*, and *lnc-776*) whose values occupied an intermediate area of the scatter plot. Therefore, while the majority of the identified lncRNAs appear as bona fide noncoding RNAs, the possibility cannot be excluded *a priori* that small open reading frames are translated into short peptides, as recently shown (34).

Analysis of muscle-specific expression. The tissue specificity of the identified lncRNAs was analyzed in different mouse body districts: skeletal muscles (gastrocnemius and tibialis) and heart and nonmuscle tissues (brain, cerebellum, and lung). Moreover, the corresponding tissues were also analyzed in the isogenic *mdx* mouse, the elective model system for muscular dystrophy (35, 36). The use of wild-type and mdx tissues provides a powerful system for comparisons of the levels of expression in mature versus regenerating fibers, which are rich in activated myoblasts. Tissues were withdrawn from 2-month-old *mdx* animals, in which a marked degeneration of muscle fibers, accompanied by intense regeneration and an increase in the number of newly differentiating myofibers, is known to occur (37). RT-PCR analysis indicated a group of at least seven species with very specific muscle-restricted expression (Fig. 2, class a). Interestingly, *lnc-405* displayed a muscle-specific splicing isoform. A subgroup of lncRNAs (i.e., *lnc-049*, *lnc-996*, and *lnc-149*) showed predominant expression in dystrophic muscles. Finally, *lnc-267* and *lnc-994* exhibited similar expression levels in *mdx* and WT mature fibers. Among the other species analyzed, the expression of *lnc-182*, *lnc-613*, and *lnc-776* was interestingly enriched in heart tissue (Fig. 2, class b). Of note, the expression of these lncRNAs was not detected in blood (see Fig. S2 in the supplemental material). Furthermore, *lnc-058*, *lnc-254*, and, to a minor extent, *lnc-023* displayed enriched expression in neuronal tissues, with *lnc-058* being upregulated in the *mdx* brain. All the remaining transcripts (Fig. 2, classes c and d) showed more ubiquitous expression which in some cases was higher in dystrophic tissues.

The RNA-seq output was then examined to locate transcription start sites (TSSs) and in turn to analyze lncRNA putative regulatory elements (Table 1). Chromatin immunoprecipitation sequencing (ChIP-seq) data sets, available from the Mouse ENCODE Project, were inspected in proliferating and differentiated C_2C_{12} cells. Orthogonal analyses allowed us to position lncRNA putative promoter regions, marked by the presence of H3K4me3 (H3 trimethyl K4), and to highlight the transcriptional activity of each locus by means of analysis of other chromatin marks (as RNA-polymerase II [RNA-Pol II] or H3K27m3 enrichment). ChIP-seq data analysis was particularly interesting in examinations of the binding of key myogenic transcription factors nearby the newly identified loci (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36024). We reasoned that engagement of muscle-type-specific transcriptional machineries was predictable for transcripts involved in the myogenic program (2, 4).

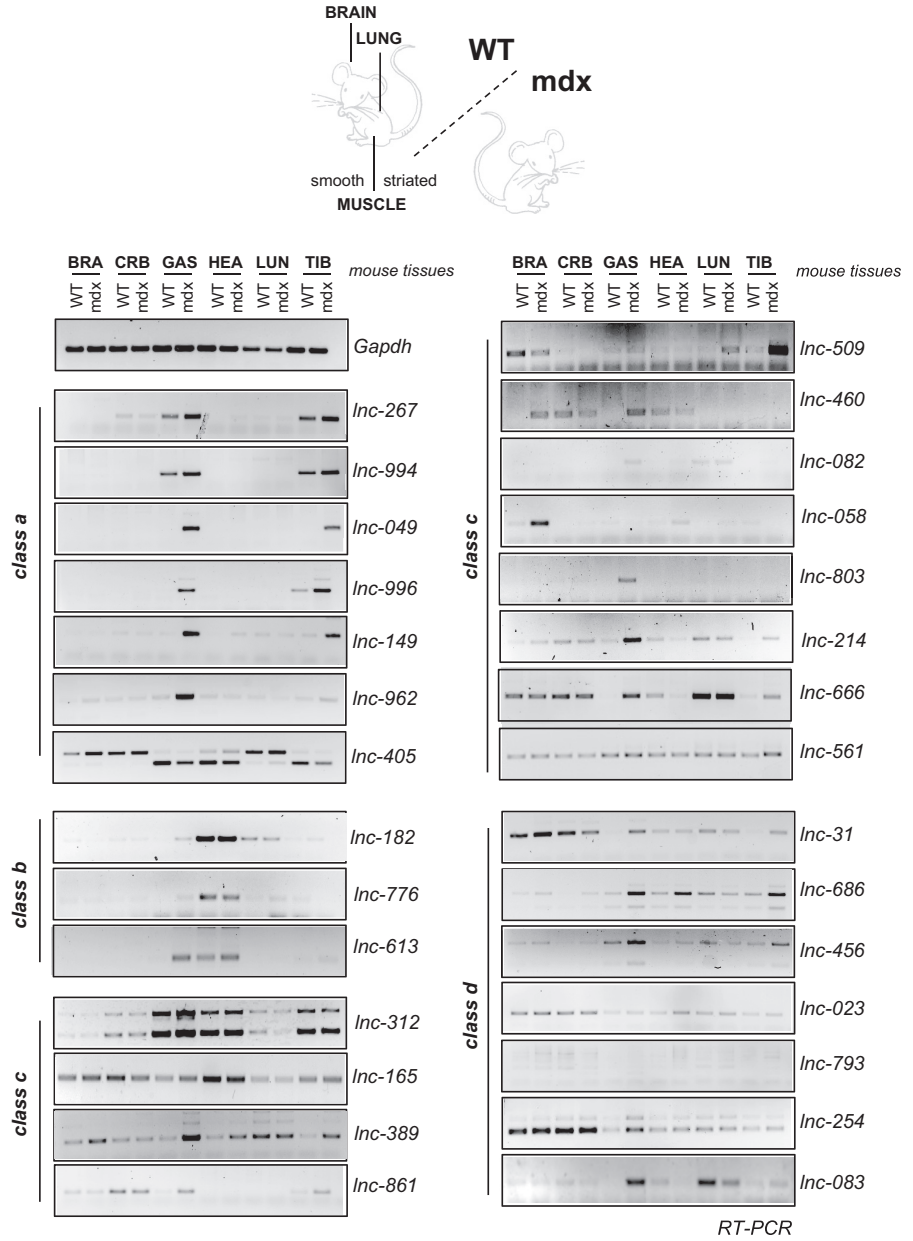

FIG 2 Tissue specificity of lncRNAs in WT and *mdx* mice. sqRT-PCRs were performed on total RNA from WT and dystrophic (*mdx*) mouse tissue samples: brain (BRA), cerebellum (CRB), gastrocnemius (GAS), heart (HEA), lung (LUN), and tibialis (TIB). lncRNAs were grouped into several classes (a to d) according to their tissue specificity: class a, upregulated and muscle-restricted lncRNAs; class b, upregulated and heart-specific lncRNAs; class c, upregulated and ubiquitously expressed lncRNAs; and class d, downregulated and ubiquitously expressed lncRNAs. *Gapdh* mRNA levels were used as a control. Analysis was performed on samples from two different individuals, and results from a representative experiment are shown.

Notably, all the upregulated species providing muscle-specific expression resulted in engagement of MYOD1 and MYOG transcription factors on regions spanning kb −5 to +1 from the putative TSS, with MYOD1 being already associated at the onset of differentiation (Table 1). MYOD1 association with target genes under proliferating (GM) conditions is a known phenomenon in myogenesis: this factor is expressed in myoblasts well before activation of its target genes and binds promoters prior to transcriptional activation. It can either act as a transcriptional repressor in partnership with histone deacetylase 1 or activate transcription upon its own acetylation (38). At variance with the group of upregulated lncRNAs, transcripts downregulated upon differentiation (Fig. 1C) showed poor expression in muscle tissues and lacked binding of myogenic factors to TSS upstream regions (Table 1). We found two exceptions: *lnc-793*, which showed a MYOD1 interaction under GM conditions, and the neural enriched *lnc-254*, which displayed a MYOG binding site close to its TSS (Table 1). Several lncRNAs belonging to classes b and c also showed MYOD1 and MYOG interaction on their upstream regions, indicating that, even if they are ubiquitously expressed, their activation in muscle cells might rely on these two factors.

TABLE 1 MYOD1 and MYOG binding sites on lncRNA putative promoters[a]

Promoter and category	MYOD1 sites in: GM SIZE_PEAK	GM DIST_TSS	DM SIZE_PEAK	DM DIST_TSS	MYOG sites in DM[b] SIZE_PEAK	MYOG sites in DM[b] DIST_TSS
Upregulated						
Class a						
lnc-267			713	299	1,455	25
			511	−364	434	−2,111
			623	−2,913	617	−2,910
lnc-994			615	−4,863	514	−4,956
lnc-049			353	15	506	32
lnc-996	936	159	1,078	148	887	200
lnc-149	572	11	649	53	460	47
	390	−591	438	−587	329	−623
	522	−2,441	702	−2,411	545	−2,408
	769	−4,150	849	−4,165	922	−4,209
lnc-962			236	−3,984	295	−939
lnc-405			524	−6,365	595	−6,486
			801	−4,503	793	−4,489
			345	−2,028	304	−2,070
			1,840	−762	1,778	−738
			677	1,166	700	1,169
Class b						
lnc-776			385	−2,882	442	−2,916
			658	−20	395	−126
lnc-613			659	226	719	238
Class c						
lnc-312	890	12	1,086	−89	1,061	−66
lnc-165			364	−1,419	502	−1,461
			409	111	431	100
					338	−573
lnc-389	564	−154	526	−77	563	−136
lnc-861			444	84	399	50
			269	−357	323	−360
lnc-509			472	−1,320		
lnc-460			363	−99		
lnc-082	385	718				
	552	−93	637	−149	528	−69
lnc-058	309	−917				
Downregulated: class d						
lnc-793	880	1,370				
lnc-254					218	−158

[a] Shown are the lncRNA species for which the occurrence of MYOD1 and MYOG binding peaks was identified by ChIP. ChIP-seq data sets from the Mouse ENCODE Project (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36024) were inspected in proliferating and differentiated cells. The base pair length (SIZE_PEAK) and distance (DIST_TSS) values of the ChIP-seq peaks from the predicted TSS are indicated. lncRNAs were grouped into upregulated and downregulated categories and further subcategorized (classes a to d) according to tissue expression: class a, upregulated and muscle-restricted lncRNAs; class b, upregulated and heart-specific lncRNAs; class c, upregulated and ubiquitously expressed lncRNAs; class d, downregulated and ubiquitously expressed lncRNAs. See also the text for details.

[b] No sites were noted in GM.

Among lncRNAs expressed in proliferating myoblasts, an interesting candidate was *lnc-31*, since it harbors precursor sequences for *miR-31*, a relevant miRNA known to play a crucial role in muscle regeneration (7, 17) and to be associated to many processes linked to cell proliferation (16). Moreover, aberrant high levels of *miR-31*, found in both human and murine dystrophic myoblasts, contribute to the pathology by compromising the normal progression of satellite cell activation to enter the differentiation program.

Subcellular localization of the newly identified lncRNAs. lncRNAs were subsequently analyzed for nucleic-cytoplasmic compartmentalization by subcellular fractionation of C_2C_{12} cells (Fig. 3; see also Fig. S3 in the supplemental material). Figure 3A shows the nucleus/cytoplasm relative abundances of lncRNAs analyzed depending on their expression category (GM or DM) (Fig. 1A and B). Since nuclear/cytoplasmic partitioning remains more difficult to obtain in terminally differentiated cells, C_2C_{12} cells were collected at day 2 of differentiation for the DM condition. Despite the correct localization of control RNAs testifying to the proper fractionation procedures (Fig. 3B), almost 12% of the species showed a bipartite distribution between the two compartments. This feature, already observed in other cases (39, 40), can account for low rates of chromatin release and cytoplasmic export for species acting in the cytoplasm; alternatively, for transcripts playing their role in the nucleus, cyto-

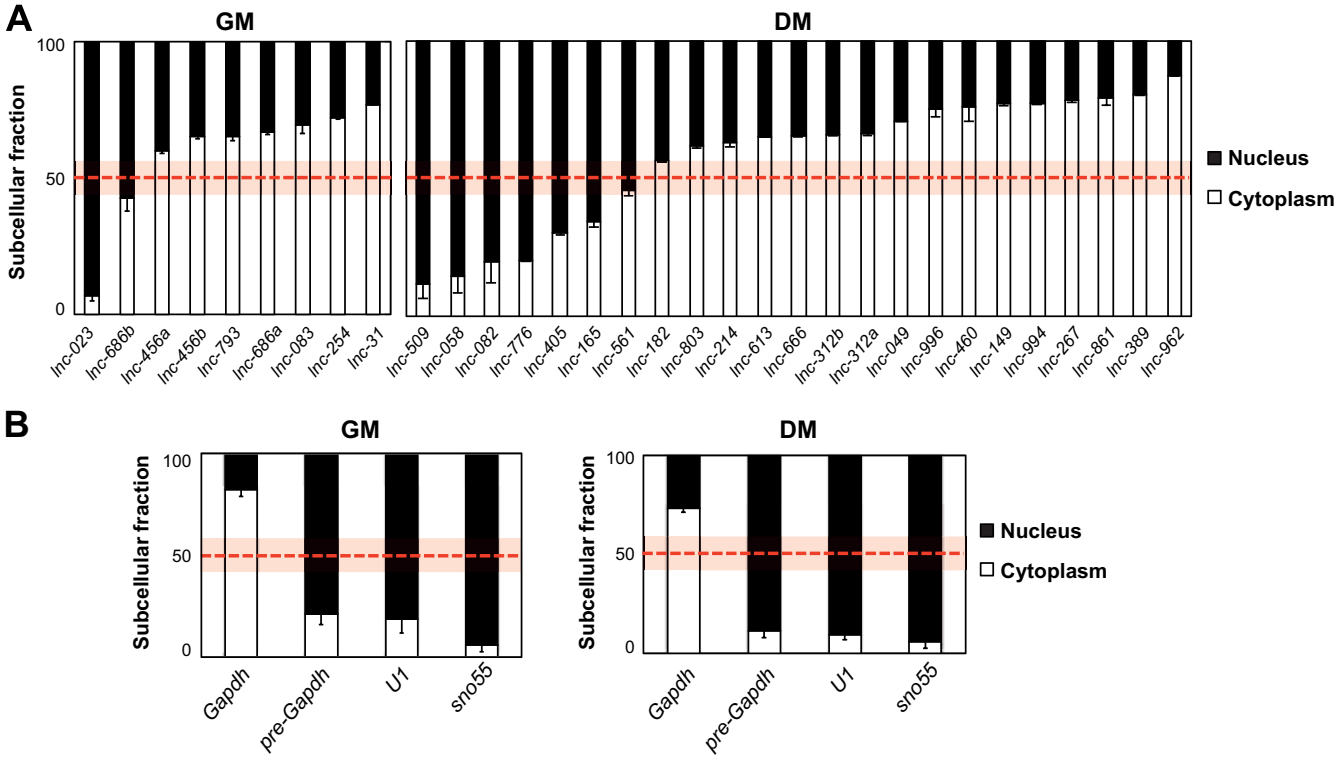

FIG 3 Subcellular distribution of the novel lncRNAs. Histograms show the quantifications of sqRT-PCR performed on RNA extracted from cytoplasmic (Cyt) and nuclear (N) fractions of myoblasts maintained under GM or DM (day 2) conditions. Results of a representative experiment are shown in Fig. S3 in the supplemental material. (A) Graphs show the cytoplasmic/nuclear partitioning of the lncRNAs, considering their expression conditions in GM or DM. The red shadow highlights the region of the graph occupied by lncRNAs with a bipartite subcellular distribution as defined by cytoplasmic/nuclear partitioning of control transcripts (see below). (B) Graphs show the quantification of the subcellular localization of control transcripts used to test the quality of the cytoplasmic/nuclear fractionation under GM or DM conditions: precursor (pre-*Gapdh*) and mature *Gapdh* mRNA, *Rnu1-2* (U1) snRNA, and *Snord55* (sno55). The red shadow defines the cytoplasmic/nuclear partition region in which no evident cytoplasmic/nuclear distribution can be assigned. Error bars represent standard deviations of data from multiple independent experiments.

plasmic localization can be due to the requirement of RNP assembly in this compartment. A third possibility of dual functions in the two compartments cannot be excluded.

***lnc-31* is a novel lncRNA associated with myogenesis.** Among the selected lncRNAs, we found a few species overlapping with miRNA genomic loci: pre-*miR125b-1* was found in the *lnc-254* unique intron, pre-*let7c2* and pre-*let7b* were found in the last exon of *lnc-023*, and pre-*miR-31* was found in the third exon of *lnc-31*. Since *miR-31* was previously shown to play an important role in muscle regeneration and to be deregulated in myopathies (7, 17, 41), *lnc-31* was selected for further studies. Moreover, a transcript overlapping *miR-31* also originates from the syntenic human HG31 genomic region at chromosome 9p21.3 (42); however, in contrast with mouse (43), whereas *miR-31* maps inside the third exon of *lnc-31* (Fig. 4A), human *miR-31* originates from the first intron of the HG31 primary transcript (see Fig. S4A in the supplemental material). Analysis of the murine *lnc-31* transcript indicated that its expression was high in proliferating C_2C_{12} cells and was strongly downregulated upon differentiation, paralleling the *lnc-31P* decrease (Fig. 4B). Notably, *miR-31* also appears to be downregulated during differentiation, even though to a lesser extent than *lnc-31*, likely due to the intrinsic higher stability of miRNAs. Interestingly, in *mdx* muscles, where more proliferating myoblasts are present due to the intense regenerative process, *lnc-31* accumulates at high levels (Fig. 2). Analogous to the murine transcript, *hsa-lnc-31* RNA is also abundant in proliferating human myoblasts and is downregulated upon differentiation (see Fig. S4B). In contrast, in human Duchenne (DMD) myoblasts, characterized by a delay in the differentiation program (18, 26, 44), *lnc-31* downregulation is less pronounced (see Fig. S4B). Taken together, these data establish a clear connection between *lnc-31* expression and myoblast proliferation, on one side, and *lnc-31* downregulation and differentiation, on the other.

Amplification with primers recognizing exonic or intronic sequences indicated that, while the primary transcript (*lnc-31P*) is exclusively nuclear, the mature *lnc-31* has a cytoplasmic localization (Fig. 4C). Treatment with siRNAs against exonic sequences resulted in efficient downregulation of mature *lnc-31* without affecting the levels of the precursor form (*lnc-31P*) or those of *miR-31* (Fig. 4D). Therefore, the RNA interference (RNAi) treatment allows the specific downregulation of the cytoplasmic *lnc-31* form and further corroborates the hypothesis that the cytoplasmic transcript escaped Drosha cleavage in the nucleus through a pathway independent of *miR-31* biogenesis. This case represents another example, similar to the case of *linc-MD1*, where two different mature molecules (a miRNA and a lncRNA) are independently produced from the same primary transcript (30).

The physiological downregulation of *lnc-31* during differentiation prompted us to analyze whether *lnc-31* controlled cell proliferation. BrdU incorporation was assessed upon RNAi treatment

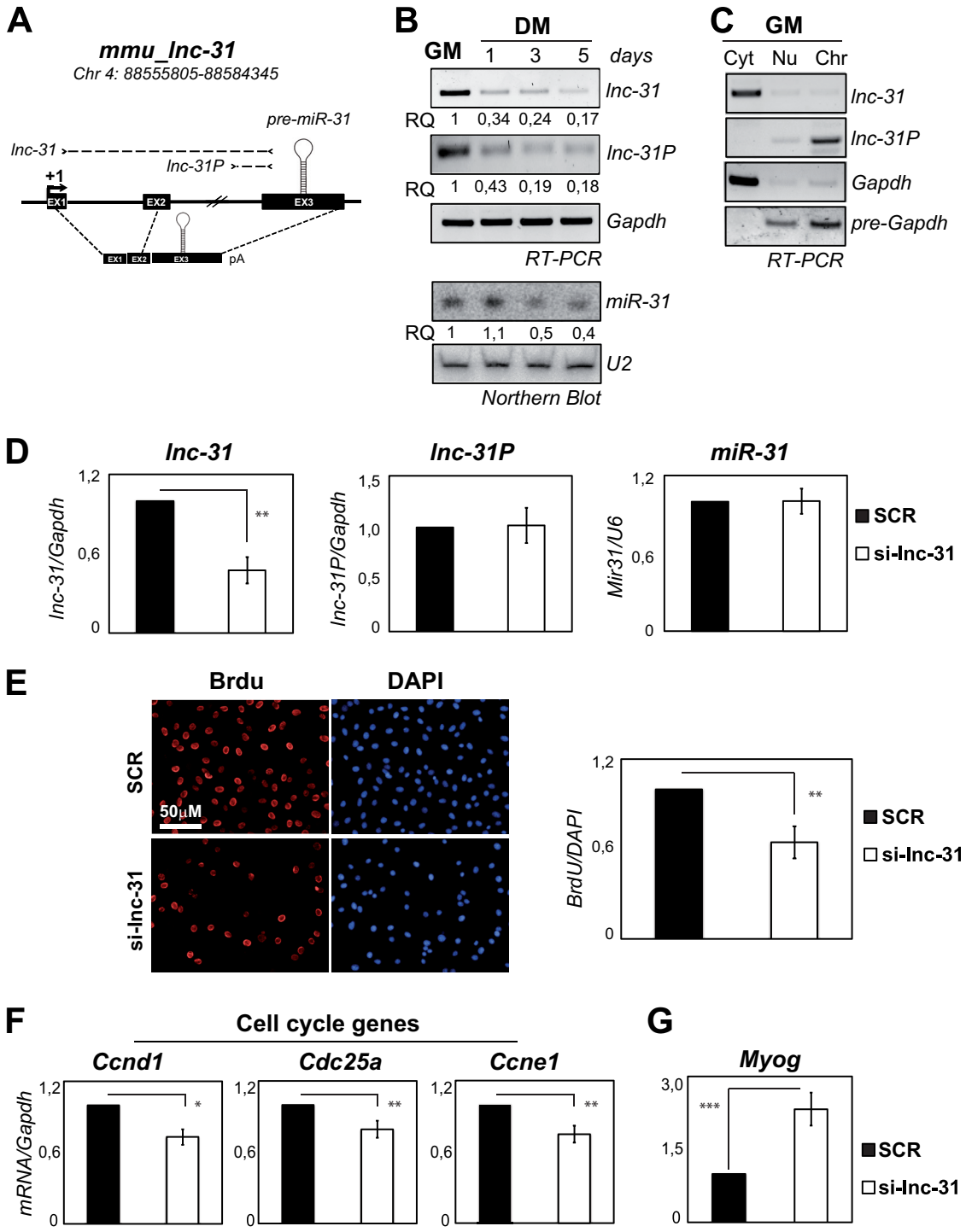

FIG 4 Murine *lnc-31* transcript analysis. (A) Schematic representation of the genomic locus encompassing the murine *lnc-31* coding region. The TSS (+1), the location of pre-*miR-31*, and the exon/intron structure are indicated together with the genomic coordinates. The arrow represents the position of the oligonucleotides used in the RT-PCR analyses; dashed lines denote the corresponding amplicons. (B) Analysis of murine *lnc-31*, *lnc-31P*, and *miR-31* expression levels determined under GM or DM (1, 3, and 5 days of differentiation) conditions. *Gapdh* mRNA and *Rnu2-10* (U2) snRNA were used as controls. Relative quantities (RQ) with respect to GM conditions are indicated below each lane; values are normalized to the amount of total *Gapdh* or U2 and represent averages of the results of three independent experiments. (C) Levels of murine *lnc-31* and of its precursor *lnc-31P* analyzed by sqRT-PCRs performed on RNA extracted from cytoplasmic (Cyt), nucleoplasmic (Nu), and chromatinic (Chr) fractions of myoblasts under GM conditions. The oligonucleotide position for RT-PCRs is shown in panel A. *Gapdh* mRNA and pre-mRNA (*pre-Gapdh*) were used as controls (see also Fig. 3B). (D) qPCR analysis of *lnc-31*, *lnc-31P*, and *miR-31* levels in C_2C_{12} cells transfected with siRNAs against *lnc-31* exonic sequences (si-lnc-31). Scramble siRNAs (SCR) were used as negative controls. *lnc-31* and *lnc-31P* expression levels were normalized to *Gapdh*, whereas *miR-31* expression levels were normalized to *Rnu6* (U6) snRNA. (E) BrdU assay performed on C_2C_{12} cells treated as described for panel D. Left panels show the BrdU and DAPI labeling, while the graph in the right panel shows the ratio between the BrdU-positive cells and the DAPI-positive cells under SCR and si-lnc-31 conditions. (F and G) Levels of *Ccnd1*, *Cdc25a*, *Ccne1*, and *Myog* mRNAs as measured by qPCR in C_2C_{12} cells treated as described for panel D. Error bars represent standard errors of results from multiple independent experiments. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ (Student's *t* test).

of C_2C_{12} cells maintained in growth medium. Compared to the results seen with cells transfected with scramble siRNAs, a 50% reduction in BrdU incorporation was observed when *lnc-31* levels were reduced to a residual 40% (Fig. 4E). In line with these results, the expression levels of cyclin D1 (*Ccnd1*), cyclin E (*Ccne1*), and *Cdc25a* cell cycle genes (45, 46) resulted in downregulation (Fig. 4F). Moreover, *lnc-31* depletion led to an increase of myogenin expression in proliferating myoblasts (Fig. 4G), which is consistent with a role of *lnc-31* in sustaining cell proliferation and in counteracting differentiation.

In order to strengthen these results, we applied RNAi against the human lnc-31 counterpart (*hsa-lnc-31*) to human primary proliferating myoblasts. Quite interestingly, this experiment provided results very similar to those in mouse. In particular, the downregulation of *hsa-lnc-31* (see Fig. S4C in the supplemental material) resulted in a 35% reduction of BrdU incorporation (see Fig. S4D and E), paralleling the decrease in the expression levels of cyclin D1 (*Ccnd1*), cyclin E (*Ccne1*), and *Cdc25a* cell cycle genes (see Fig. S4F). Finally, in agreement with the mouse results, the levels of myogenin that resulted increased (see Fig. S4G), indicating an overall effect of *hsa-lnc-31* in sustaining cell proliferation and counteracting differentiation.

Analysis of the expression levels of *lnc-31* in *mdx* animals, where intense regeneration occurs through myoblast activation (37), and in human DMD myoblasts, which have a delayed differentiation phenotype (18, 26, 44), further supports the notion that *lnc-31* inversely correlates with the ability of myogenic cells to progress into the differentiation program.

Taken together, these data show that *lnc-31* is able, independently from *miR-31*, to control the differentiation commitment of precursor myoblasts and indicate that its function is conserved in evolution despite the poor sequence conservation seen with the human counterpart.

ACKNOWLEDGMENTS

We thank A. Musarò and C. Nicoletti for animal facilities, A. Lund and M. Montes for *hsa-lnc-31* siRNAs, O. Sthandier and M. Marchioni for technical support, and J. Hughes for helpful discussion. We are grateful to the NGS platform of La Sapienza University and Bio-Fab research for their technical expertise and commitment.

This work was partially supported by grants from ERC (ERC-2013-AdG340172-MUNCODD), Telethon (GGP11149), the Epigen-Epigenomics Flagship Project, Parent Project Italia, and PRIN to I.B. and KAUST and PRIN to A.T.

We declare that we have no conflicts of interest.

REFERENCES

1. **Edmondson DG, Olson EN.** 1989. A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev **3:**628–640. http://dx.doi.org/10.1101/gad.3.5.628.
2. **Moncaut N, Rigby PWJ, Carvajal JJ.** 2013. Dial M(RF) for myogenesis. FEBS J **280:**3980–3990. http://dx.doi.org/10.1111/febs.12379.
3. **Perry RL, Rudnick MA.** 2000. Molecular mechanisms regulating myogenic determination and differentiation. Front Biosci **5:**D750–D767. http://dx.doi.org/10.2741/Perry.
4. **Sartorelli V, Caretti G.** 2005. Mechanisms underlying the transcriptional regulation of skeletal myogenesis. Curr Opin Genet Dev **15:**528–535. http://dx.doi.org/10.1016/j.gde.2005.04.015.
5. **Fatica A, Bozzoni I.** 2014. Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet **15:**7–21. http://dx.doi.org/10.1038/nrg3606.
6. **van Rooij E, Liu N, Olson EN.** 2008. MicroRNAs flex their muscles. Trends Genet **24:**159–166. http://dx.doi.org/10.1016/j.tig.2008.01.007.
7. **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:**136–141. http://dx.doi.org/10.1038/embor.2010.208.
8. **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:**341–351. http://dx.doi.org/10.1016/j.cmet.2010.07.008.
9. **Chen J-F, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang D-Z.** 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet **38:**228–233. http://dx.doi.org/10.1038/ng1725.
10. **Eisenberg I, Eran A, Nishino I, Moggio M, Lamperti C, Amato AA, Lidov HG, Kang PB, North KN, Mitrani-Rosenbaum S, Flanigan KM, Neely LA, Whitney D, Beggs AH, Kohane IS, Kunkel LM.** 2007. Distinctive patterns of microRNA expression in primary muscular disorders. Proc Natl Acad Sci U S A **104:**17016–17021. http://dx.doi.org/10.1073/pnas.0708115104.
11. **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:**358–369. http://dx.doi.org/10.1016/j.cell.2011.09.028.
12. **Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA.** 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell **152:**570–583. http://dx.doi.org/10.1016/j.cell.2013.01.003.
13. **Wang J, Gong C, Maquat LE.** 2013. Control of myogenesis by rodent SINE-containing lncRNAs. Genes Dev **27:**793–804. http://dx.doi.org/10.1101/gad.212639.112.
14. **Rinn JL, Chang HY.** 2012. Genome regulation by long noncoding RNAs. Annu Rev Biochem **81:**145–166. http://dx.doi.org/10.1146/annurev-biochem-051410-092902.
15. **Twayana S, Legnini I, Cesana M, Cacchiarelli D, Morlando M, Bozzoni I.** 2013. Biogenesis and function of non-coding RNAs in muscle differentiation and in Duchenne muscular dystrophy. Biochem Soc Trans **41:**844–849. http://dx.doi.org/10.1042/BST20120353.
16. **Valastyan S, Weinberg RA.** 2010. miR-31: a crucial overseer of tumor metastasis and other emerging roles. Cell Cycle **9:**2124–2129. http://dx.doi.org/10.4161/cc.9.11.11843.
17. **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:**118–126. http://dx.doi.org/10.1016/j.stem.2012.03.011.
18. **Cazzella V, Martone J, Pinnarò C, Santini T, Twayana SS, Sthandier O, D'Amico A, Ricotti V, Bertini E, Muntoni F, Bozzoni I.** 2012. Exon 45 skipping through U1-snRNA antisense molecules recovers the Dys-nNOS pathway and muscle differentiation in human DMD myoblasts. Mol Ther **20:**2134–2142. http://dx.doi.org/10.1038/mt.2012.178.
19. **Trapnell C, Pachter L, Salzberg SL.** 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25:**1105–1111. http://dx.doi.org/10.1093/bioinformatics/btp120.
20. **Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L.** 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc **7:**562–578. http://dx.doi.org/10.1038/nprot.2012.016.
21. **Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L.** 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28:**511–515. http://dx.doi.org/10.1038/nbt.1621.
22. **Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L.** 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol **31:**46–53. http://dx.doi.org/10.1038/nbt.2450.
23. **Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R.** 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene struc-

ture, evolution, and expression. Genome Res **22:**1775–1789. http://dx.doi.org/10.1101/gr.132159.111.

24. **Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, et al.** 2012. Landscape of transcription in human cells. Nature **489:**101–108. http://dx.doi.org/10.1038/nature11233.
25. **Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ.** 2008. Primary microRNA transcripts are processed co-transcriptionally. Nat Struct Mol Biol. **15:**902–909. http://dx.doi.org/10.1038/nsmb.1475.
26. **Incitti T, De Angelis FG, Cazzella V, Sthandier O, Pinnarò C, Legnini I, Bozzoni I.** 2010. Exon skipping and Duchenne muscular dystrophy therapy: selection of the most active U1 snRNA antisense able to induce dystrophin exon 51 skipping. Mol Ther **18:**1675–1682. http://dx.doi.org/10.1038/mt.2010.123.
27. **Chargé SBP, Rudnicki MA.** 2004. Cellular and molecular regulation of muscle regeneration. Physiol Rev **84:**209–238. http://dx.doi.org/10.1152/physrev.00019.2003.
28. **Wagers AJ, Conboy IM.** 2005. Cellular and molecular signatures of muscle regeneration: current concepts and controversies in adult myogenesis. Cell **122:**659–667. http://dx.doi.org/10.1016/j.cell.2005.08.021.
29. **D'Andrea D, Grassi L, Mazzapioda M, Tramontano A.** 2013. FIDEA: a server for the functional interpretation of differential expression analysis. Nucleic Acids Res **41**(Web Server issue):W84–W88. http://dx.doi.org/10.1093/nar/gkt516.
30. **Legnini I, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I.** 2014. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. Mol Cell **53:**506–514. http://dx.doi.org/10.1016/j.molcel.2013.12.012.
31. **Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G.** 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res **35**(Web Server issue):W345–W349. http://dx.doi.org/10.1093/nar/gkm391.
32. **Arrial RT, Togawa RC, Brigido Mde M.** 2009. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus Paracoccidioides brasiliensis. BMC Bioinformatics **10:**239. http://dx.doi.org/10.1186/1471-2105-10-239.
33. **Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W.** 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res **41:**e74. http://dx.doi.org/10.1093/nar/gkt006.
34. **Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, Giraldez AJ.** 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J **33:**981–993. http://dx.doi.org/10.1002/embj.201488411.
35. **Dunckley MG, Manoharan M, Villiet P, Eperon IC, Dickson G.** 1998. Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. Hum Mol Genet **7:**1083–1090. http://dx.doi.org/10.1093/hmg/7.7.1083.
36. **Willmann R, Possekel S, Dubach-Powell J, Meier T, Ruegg MA.** 2009. Mammalian animal models for Duchenne muscular dystrophy. Neuromuscul Disord **19:**241–249. http://dx.doi.org/10.1016/j.nmd.2008.11.015.
37. **Muntoni F, Mateddu A, Marchei F, Clerk A, Serra G.** 1993. Muscular weakness in the mdx mouse. J Neurol Sci **120:**71–77. http://dx.doi.org/10.1016/0022-510X(93)90027-V.
38. **Mal A, Harter ML.** 2003. MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. Proc Natl Acad Sci U S A **100:**1735–1739. http://dx.doi.org/10.1073/pnas.0437843100.
39. **Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR.** 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science **316:**1484–1488. http://dx.doi.org/10.1126/science.1138341.
40. **van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, Macinnes AW, Cuppen E, Simonis M.** 2014. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. Genome Biol **15:**R6. http://dx.doi.org/10.1186/gb-2014-15-1-r6.
41. **Roberts TC, Blomberg KEM, McClorey G, El Andaloussi S, Godfrey C, Betts C, Coursindel T, Gait MJ, Smith CIE, Wood MJA.** 2012. Expression analysis in multiple muscle groups and serum reveals complexity in the microRNA transcriptome of the mdx mouse with implications for therapy. Mol Ther Nucleic Acids **1:**e39. http://dx.doi.org/10.1038/mtna.2012.26.
42. **Augoff K, McCue B, Plow EF, Sossey-Alaoui K.** 2012. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. Mol Cancer **11:**5. http://dx.doi.org/10.1186/1476-4598-11-5.
43. **Sun F, Wang J, Pan Q, Yu Y, Zhang Y, Wan Y, Wang J, Li X, Hong A.** 2009. Characterization of function and regulation of miR-24-1 and miR-31. Biochem Biophys Res Commun **380:**660–665. http://dx.doi.org/10.1016/j.bbrc.2009.01.161.
44. **Delaporte C, Dehaupas M, Fardeau M.** 1984. Comparison between the growth pattern of cell cultures from normal and Duchenne dystrophy muscle. J Neurol Sci **64:**149–160. http://dx.doi.org/10.1016/0022-510X(84)90033-9.
45. **Jahn L, Sadoshima J, Izumo S.** 1994. Cyclins and cyclin-dependent kinases are differentially regulated during terminal differentiation of C2C12 muscle cells. Exp Cell Res **212:**297–307. http://dx.doi.org/10.1006/excr.1994.1147.
46. **Sarkar S, Dey BK, Dutta A.** 2010. MiR-322/424 and -503 are induced during muscle differentiation and promote cell cycle quiescence and differentiation by downregulation of Cdc25A. Mol Biol Cell **21:**2138–2149. http://dx.doi.org/10.1091/mbc.E10-01-0062.