

RESEARCH ARTICLE

Open Access



Genome-wide identification and characterization of long non-coding RNAs in developmental skeletal muscle of fetal goat

Siyuan Zhan[†], Yao Dong[†], Wei Zhao, Jiazhong Guo, Tao Zhong, Linjie Wang, Li Li^{*} and Hongping Zhang^{*}

Abstract

Background: Long non-coding RNAs (lncRNAs) have been studied extensively over the past few years. Large numbers of lncRNAs have been identified in mouse, rat, and human, and some of them have been shown to play important roles in muscle development and myogenesis. However, there are few reports on the characterization of lncRNAs covering all the development stages of skeletal muscle in livestock.

Results: RNA libraries constructed from developing *longissimus dorsi* muscle of fetal (45, 60, and 105 days of gestation) and postnatal (3 days after birth) goat (*Capra hircus*) were sequenced. A total of 1,034,049,894 clean reads were generated. Among them, 3981 lncRNA transcripts corresponding to 2739 lncRNA genes were identified, including 3515 intergenic lncRNAs and 466 anti-sense lncRNAs. Notably, in pairwise comparisons between the libraries of skeletal muscle at the different development stages, a total of 577 transcripts were differentially expressed ($P < 0.05$) which were validated by qPCR using randomly selected six lncRNA genes. The identified goat lncRNAs shared some characteristics, such as fewer exons and shorter length, with the lncRNAs in other mammals. We also found 1153 lncRNAs genes were neighbored 1455 protein-coding genes (<10 kb upstream and downstream) and functionally enriched in transcriptional regulation and development-related processes, indicating they may be in *cis*-regulatory relationships. Additionally, Pearson's correlation coefficients of co-expression levels suggested 1737 lncRNAs and 19,422 mRNAs were possibly in *trans*-regulatory relationships ($r > 0.95$ or $r < -0.95$). These co-expressed mRNAs were enriched in development-related biological processes such as muscle system processes, regulation of cell growth, muscle cell development, regulation of transcription, and embryonic morphogenesis.

Conclusions: This study provides a catalog of goat muscle-related lncRNAs, and will contribute to a fuller understanding of the molecular mechanism underpinning muscle development in mammals.

Keywords: Muscle development, lncRNA, Goat, Transcriptome, *cis*-acting, *trans*-acting, Differential expression

Background

Genome-wide transcriptional studies have revealed that large regions of eukaryotic genomes are transcribed into a heterogeneous population of non-coding RNAs (ncRNAs). Generally, ncRNAs shorter than 200 nucleotides are usually described as small/short ncRNA, such as microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), small

interfering RNAs (siRNAs), and classical ncRNAs such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nucleolar RNAs (snoRNAs), whereas ncRNAs longer than 200 nucleotides are described as long ncRNAs (lncRNAs). In the past few years, an increasing number of lncRNAs have been discovered in mammal, including *Homo sapiens* [1, 2], *Mus musculus* [3–7], *Bos taurus* [8, 9], *Sus scrofa* [10–13], and *Ovis aries* [14]. And accordingly unveiled that lncRNAs play critical roles in biological processes like transcriptional regulation [15–17], epigenetic modification [18–20], development

* Correspondence: lilyzh002@sohu.com; zhp@sicau.edu.cn

[†]Equal contributors

Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu 611130, China



[21–23], cell differentiation [24–26], as well as in some diseases [27–29].

As important economic animals worldwide, domestic goats (*Capra hircus*) are raised mainly for meat production. Thus, unveiling the molecular mechanisms underneath skeletal muscle formation and development is of vital interest. Muscle development is a complex process that requires the concerted expression and interaction of multiple factors [30]. Several recent studies have indicated that lncRNAs play crucial roles in myogenic differentiation and myogenesis [31–35]. Nevertheless, currently the majority strategy for exploring molecular mechanisms underlying skeletal muscle growth and development in mammals [36–38] is detecting the expression and functions of coding genes like the MRF (myogenic regulatory factor) [39, 40], MEF2 (myocyte enhancer factor-2) [41, 42] families, and the paired box proteins [43], though high-throughput sequencing technologies is also employed to profile expression of miRNA and mRNA expression in goat [44, 45]. Therefore, information about skeletal muscle development-related lncRNAs is still limited especially in goats.

Here, we report the systematic identification and characterization of lncRNAs in fetal and postnatal goat skeletal muscle using an Illumina HiSeq 2500 platform. A total of 3981 lncRNA transcripts were identified and 577 of these transcripts were significantly differentially expressed in pairwise comparisons between RNA libraries of skeletal muscle at the different development stages. To the best of our knowledge, no other report on muscle lncRNAs and their biological functions in goat is currently available. Our results will provide a useful resource for better understanding the regulatory functions of lncRNAs in goat and for annotating the goat genome, as well as contribute to better comprehending skeletal muscle development in mammals.

Results

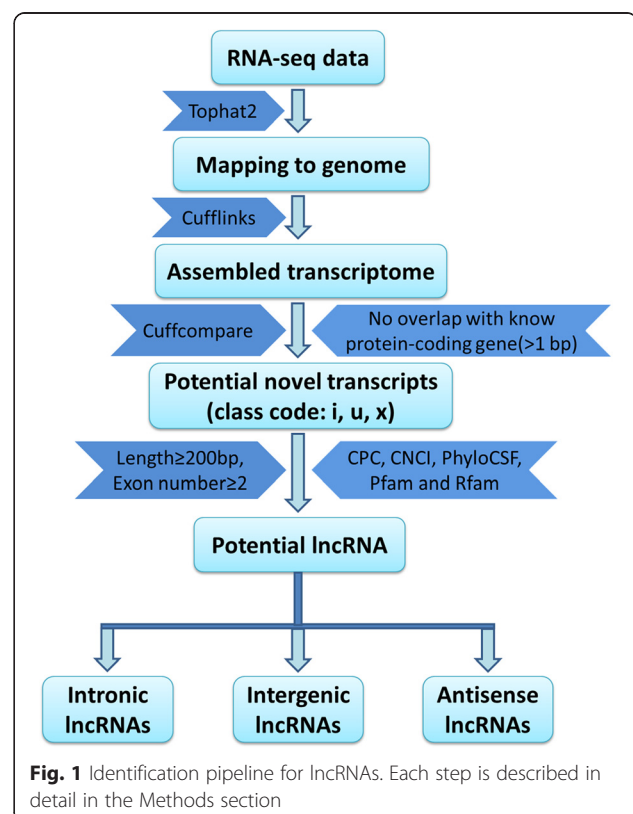
Overview of RNA sequencing (RNA-seq)

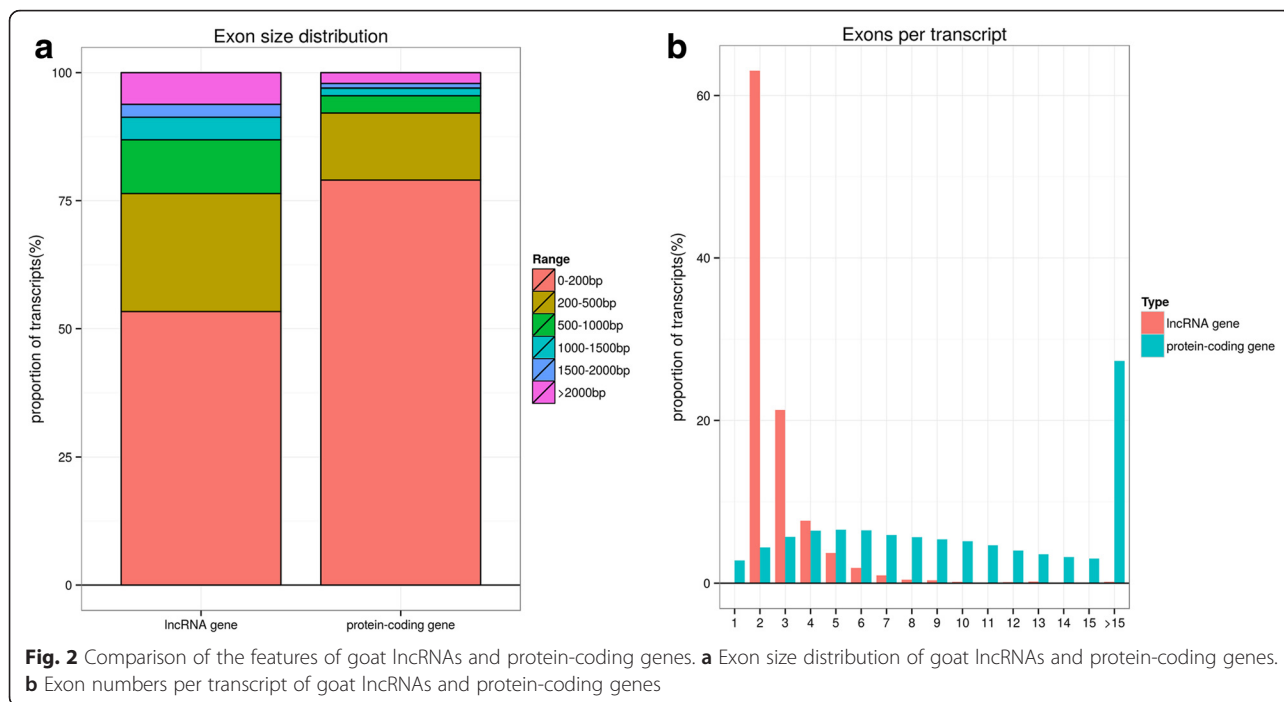
To identify lncRNAs expressed in goat skeletal muscle development, we constructed 11 cDNA libraries (E45-1, E45-2, E45-3, E60-1, E60-2, E105-1, E105-2, E105-3, B3-1, B3-2, B3-3) from goat *longissimus dorsi* muscle samples at four developmental stages: three gestation stages at 45, 60, and 105 days of gestation (E45, E60, and E105), and one postnatal stage (B3). Three biological replicates for E45, E105, and B3, and two biological replicates for E60 were used. The libraries were sequenced using an Illumina HiSeq 2500 platform and 125 bp paired-end reads were generated. A total of 1,052,994,828 raw reads were generated in all 11 libraries. After discarding adaptor sequences and low-quality reads, we obtained 1,034,049,894 clean reads. The percentage of clean reads in each library ranged from 97.90 to 98.53 % (for details of

the sequencing results see Additional file 1). We mapped the clean reads to the goat reference genome sequence (CHIR_1.0, NCBI). Approximately 75.20–86.60 % of the clean reads in all the libraries were mapped to the goat reference genome (Additional file 1). The mapped sequences in each library were assembled and a total of 56,710 unique assembled transcripts were obtained.

Identification of lncRNAs in goat skeletal muscle

We developed a highly stringent filtering pipeline to discard transcripts that did not have all the characteristics of lncRNAs (Fig. 1). Our pipeline yielded 3981 lncRNA transcripts, including 3515 intergenic lncRNAs (88.29 %) and 466 anti-sense lncRNAs (11.71 %) (Additional file 2). These transcripts corresponded to 2739 lncRNA genes, an average of 1.5 transcripts per lncRNA locus. We found that the lncRNA transcripts were distributed in all chromosomes except the Y chromosome (Additional file 3). The Illumina RNA-seq also produced 24,383 protein-coding transcripts with an average length of 1978 bp and 8.4 exons, which was longer than the lncRNA genes, which averaged 1296 bp in length and 2.4 exons. However, the exon size in the protein-coding genes was smaller than the exon size in the lncRNA genes (most of protein-coding genes were within 200 bp) (Fig. 2a). We also found that protein transcripts with two and three exons accounted for 10.6 % of all the protein-coding genes,





which was much lower than the percentage of lncRNA genes with two and three exons (Fig. 2b).

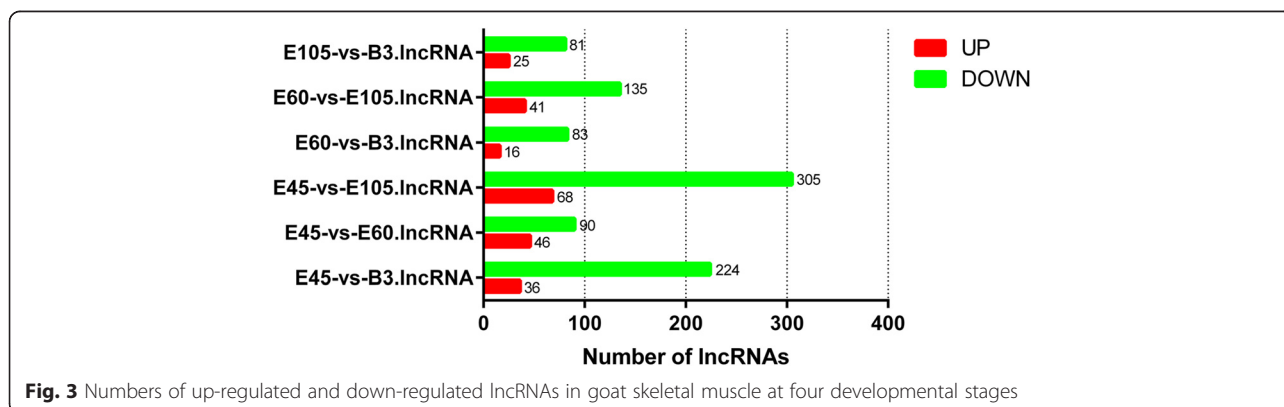
Identification of differentially expressed lncRNAs

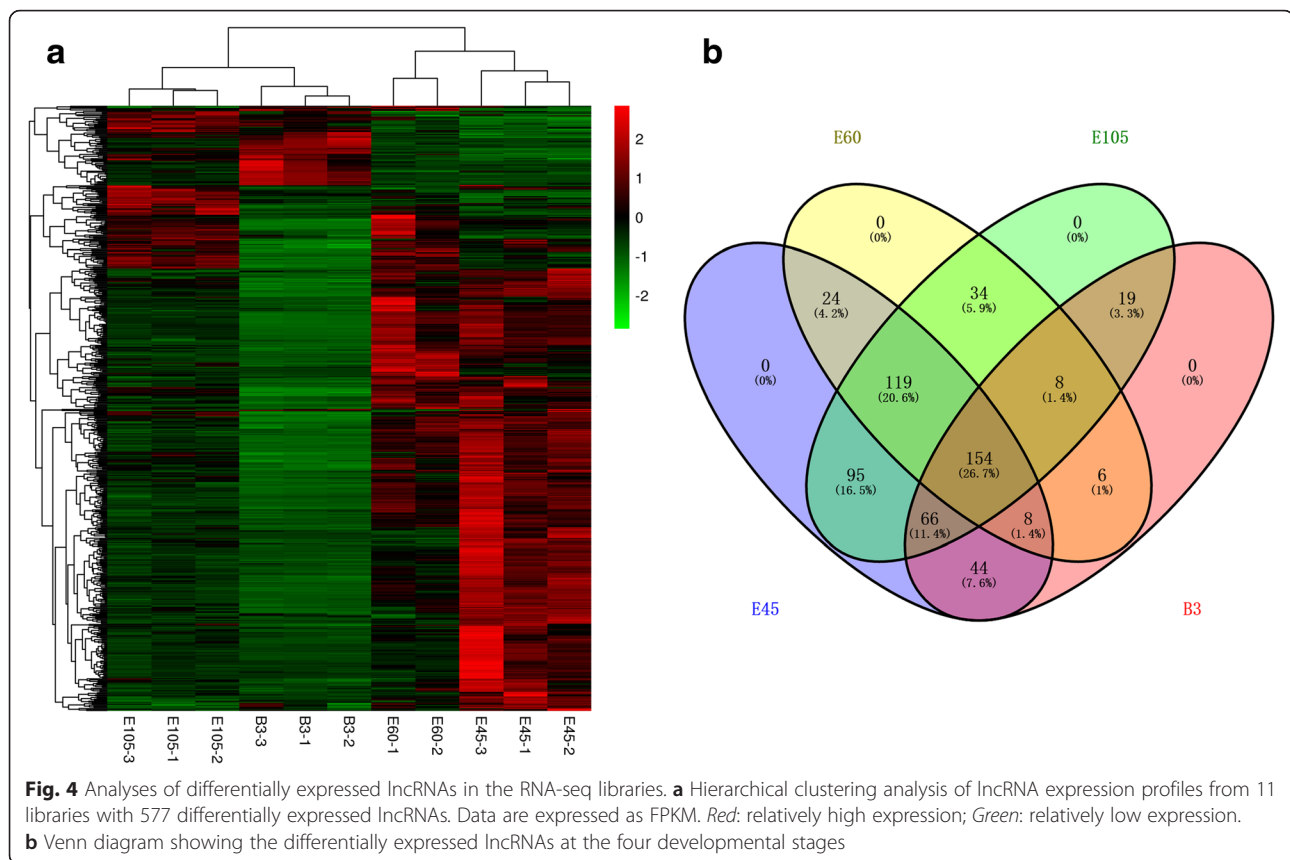
The expression levels of the lncRNA transcripts were estimated by FPKM (fragments per kilo-base of exon per million fragments mapped) using Cuffdiff. We identified 577 lncRNA transcripts that were differentially expressed during muscle development (Fig. 3 and Additional file 4); the number of down-regulated lncRNAs was higher than the number of up-regulated lncRNAs during development. The expression patterns of differentially expressed lncRNAs were measured by systematic cluster analysis, to explore the similarities and to compare the relationships between the different libraries (Fig. 4a and additional file 5). The replicates for each developmental stage clustered

together, and E45 and E60 formed one group and E105 and B3 formed another group. To further analyze the interactions among the differentially expressed lncRNAs, we constructed a Venn diagram using the 510, 353, 495, and 435 lncRNAs that were differentially expressed in E45, E60, E105, and B3, respectively. We did not detect any stage-specific differentially expressed lncRNAs among the four developmental stages, but we identified 154 differentially expressed lncRNAs that were detected in all four developmental stages (Fig. 4b).

Enrichment analysis of nearest neighbor genes of the lncRNAs

To investigate the possible functions of the lncRNAs, we predicted the potential targets of lncRNAs in *cis*-regulatory relationships. We searched for protein-coding genes 10-kb





upstream and downstream of all the identified lncRNAs. We found 1153 lncRNAs that were transcribed close to (<10 kb) 1455 protein-coding neighbors (Additional file 6). Gene Ontology (GO) [46] analysis of the *cis* lncRNA targets was performed to explore their functions. We found 88 GO terms that were significantly enriched ($P < 0.05$), and 12 of these terms were associated with regulation of gene expression. For example, the top 10 enriched terms included nucleotide binding, regulation of RNA metabolic process, DNA-dependent regulation of transcription, transcription regulator activity, and transcription factor activity (Additional file 7). These results suggest that one of the principal roles of lncRNAs may be transcriptional regulation of gene expression. Interestingly, we also found genes, including *RBP4*, *PLN*, *MYLK2*, *RARA*, *CACNB4*, *NR2F2*, *CDK5* and *PITXI*, that were annotated with muscle development-related GO term, striated muscle tissue development (GO:0014706). These results suggest that muscle development may be regulated by the action of lncRNAs on these neighboring protein-coding genes. Pathway analysis [47] showed that the 1547 candidate *cis* target genes were enriched in 252 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, several of which were related to muscle development such as insulin signaling pathway, MAPK signaling pathway, TGF-beta signaling pathway, and PPAR signaling pathway (Additional file 7).

Enrichment analysis of co-expressed genes of lncRNAs

We also predicted the potential targets of lncRNAs in *trans*-regulatory relationships using co-expression analysis. A total of 288,020 interaction relationships (285,161 positive and 2859 negative correlations) were detected between 1747 lncRNA transcripts and 19,846 protein-coding transcripts that corresponded to 7718 protein-coding genes in the goat reference genome (Additional file 8). Functional analysis showed that the co-expressed genes were enriched in 446 GO terms (253 under biological process, 91 GO under cellular component, and 102 under molecular function) that encompassed a variety of biological processes (Additional file 9). Importantly, some of the terms were muscle development-related terms, including muscle cell development (GO:0055001), striated muscle cell development (GO:0055002), muscle contraction (GO:0006936), and muscle system process (GO:0003012). In addition, the co-expressed genes were enriched in 285 KEGG pathways, several of which were related to muscle development, including TGF-beta signaling, MAPK signaling, and PPAR signaling pathways (Additional file 9). These findings indicate that lncRNAs also regulate *trans* target genes.

Validation of differentially expressed lncRNAs

We randomly selected six differentially expressed lncRNAs and examined their expression patterns at four

developmental stages by qPCR. The results confirmed that the six lncRNAs were expressed at all four development stages (Fig. 5) and showed differential expression at different stages. In addition, the qPCR confirmed that the expression patterns of the six lncRNAs were consistent with their expression levels calculated from the RNA-seq data. All our results show that our pipeline was highly strict in identifying putative lncRNAs, and indicate that most of the identified lncRNAs were truly expressed in vivo.

Discussion

The identification and characterization of goat lncRNAs, particularly in fetal skeletal muscle development, have been very limited compared with lncRNAs in human [2, 48] and other model organisms [3, 49]. In goat skeletal muscle, the main focus has been on genes and miRNAs rather than on lncRNAs [44, 45, 50, 51]. In the present study, we identified a total of 3981 multiple exon lncRNAs in fetal and postnatal goat skeletal muscle. To the best of our knowledge, this is the first report to systematically identify lncRNAs from RNA-seq data during goat skeletal muscle development.

Non-coding and protein-coding genes are distinguished by their potential coding capability. In this study, we developed a highly stringent filtering pipeline (Fig. 1) to minimize the selection of false positive lncRNAs, which aimed to remove transcripts with evidence of protein-coding potential. We identified 3981 putative lncRNAs with high confidence across four development stages of goat skeletal muscle. In agreement with similar studies on different organisms, the identified putative lncRNAs had fewer exon numbers, shorter transcript lengths, and lower expression levels than protein-coding genes [11, 48, 49]. The number of putative lncRNAs detected in this study was more than that reported in previous studies in cattle and goat [9, 52]. Six randomly selected differentially expressed lncRNA transcripts were validated using qPCR, and the results were consistent with the results from the RNA-seq data. Together, these results confirmed that the identified lncRNAs were of high quality.

lncRNAs are a group of endogenous RNAs that function as regulators of gene expression, and are involved in developmental and physiological processes [23, 53, 54]. We detected 577 putative lncRNAs that were differentially expressed in pairwise comparisons between goat skeletal

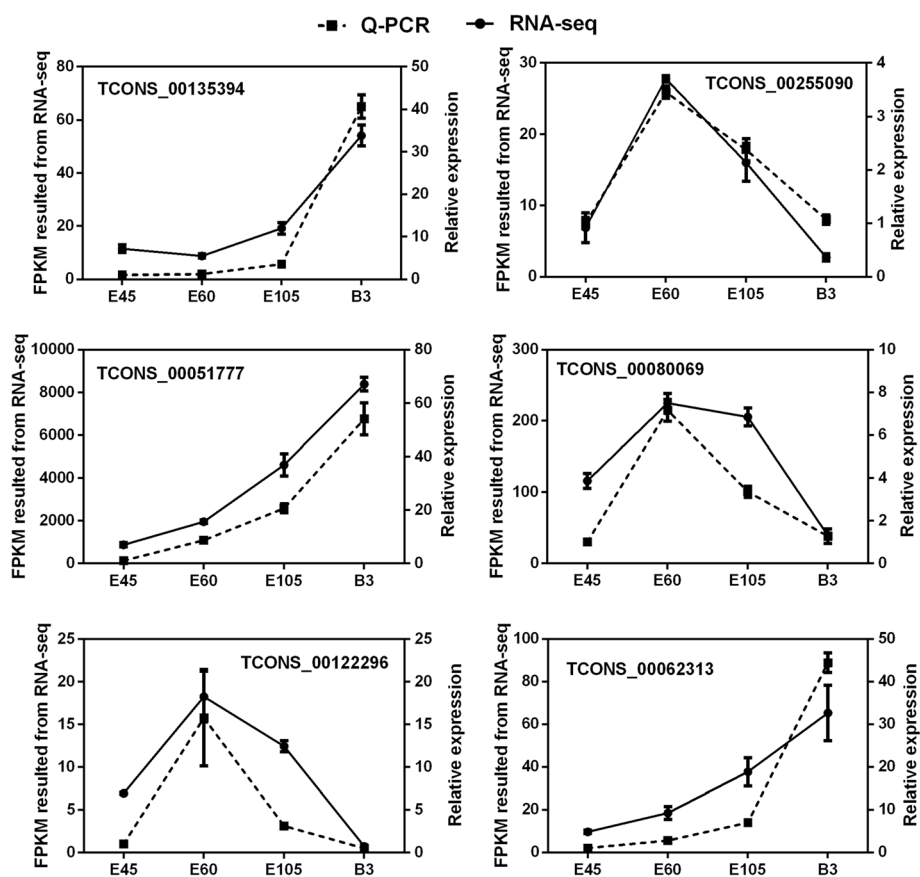


Fig. 5 Validation of six differentially expressed lncRNAs by qPCR. Data are the mean \pm SEM

muscle at the different development stages. These lncRNAs may have specific biological roles in early muscle development in fetal goat. Skeletal muscle development from the fetal to the adult stage consists of a series of exquisitely regulated and orchestrated changes in the expression of many genes [55]. In recent years, the roles of some lncRNAs in muscle biology have been reported. For example, the long intergenic ncRNA muscle differentiation linc-MD1 was the first muscle-specific lncRNA to be identified [24]. Linc-MD1 is required for appropriate muscle differentiation, at least in part because it regulates the levels of myocyte enhancer factor 2C (MEF2C) and mastermind-like protein 1 (MAML1) by sponging endogenous miR-133 and miR-135 in the cytoplasm of muscle cells [24]. In addition, the substantial disintegration of linc-MD1 in primary myoblasts of patients with Duchenne muscular dystrophy suggests that it is likely involved in the pathogenesis of this muscle disorder [24]. Another study revealed that the lncRNA, lncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation [56]. Therefore, the differentially expressed lncRNAs reported here can be considered as important novel regulators of goat skeletal muscle biology.

Unlike miRNAs or proteins, the functions of ncRNAs cannot currently be inferred from their sequence or structure; therefore, in this study, we predicted the potential function of the detected lncRNAs using *cis* and *trans* methods. The *cis* nature of a lncRNA refers to its ability to act on a neighboring gene on the same allele from which it is transcribed; thus, this type of lncRNA commonly forms a feedback loop that regulates itself and its neighboring genes.

In the *cis* prediction, we searched for coding genes 10-kb upstream and downstream of all the identified lncRNAs. GO and KEGG analyses of the neighboring protein-coding genes revealed that major enriched pathways were associated with transition metal ion binding, nucleotide binding, zinc ion binding, regulation of RNA metabolic process, regulation of transcription, and transcription regulator activity. These results indicate the possible role of lncRNAs in transcriptional regulation of gene expression. Interestingly, we found some of the *cis* target protein-coding genes were involved in skeletal muscle tissue development (e.g., *MYLK2*, *NR2F2*, *CDK5* and *PITX1*) (Additional file 6), implying that the corresponding lncRNAs play regulatory roles in skeletal muscle development. Several recent studies also indicated that lncRNAs were involved in *cis*-regulatory activity in muscle development; for example, the lncRNA Dum (developmental pluripotency-associated 2 (Dppa2) upstream binding muscle lncRNA) was identified in skeletal myoblast cells [31]. Dum promotes myoblast differentiation and damage-induced muscle regeneration by silencing its neighboring gene, Dppa2, in *cis* through recruiting Dnmt1, Dnmt3a, and Dnmt3b [31]. Similarly, a

ChIP-seq study of the Yin Yang 1 (YY1) transcription factor, an important component of the PcG complex that negatively regulates myogenesis, identified a number of lncRNAs regulated by YY1 (YY1-associated muscle lncRNAs or Yams) [57]. Among the Yams, Yam-1 displayed a *cis* effect on the expression of neighboring genes, including one that encodes miR-715, which targets and represses *Wnt7b* in skeletal muscle.

Many lncRNAs can also function in *trans* mode to target gene loci distant from where the lncRNAs are transcribed [58]. In the co-expression analysis, we detected 1747 lncRNA transcripts that were related to protein-coding genes based on the expression correlation coefficient ($r > 0.95$ or < -0.95). GO enrichment analysis found that the enriched GO terms referred mainly to development process, transcriptional regulation, and biosynthetic process. Furthermore, a cluster of lncRNAs in the co-expression analysis often targeted protein-coding genes that were expressed specifically in muscle and were involved in muscle development (e.g., *TNNT1*, *TNNT3*, *MYH1*, *MYH2*, *MyoG*, and *MYL3*). This is an interesting observation, which indicates the functional complexity of lncRNAs and is worth investigating further. Mousavi et al. [59] found two lncRNAs in two enhancer regions of the *MyoD* gene that they named ^{DDR}RNA and ^{CE}RNA, where DRR indicates distal regulatory regions and CE indicates core enhancer. The study showed that ^{CE}RNA facilitated the occupancy of RNA polymerase II in *cis* by increasing chromatin accessibility, stimulating the expression of *MyoD*, while ^{DDR}RNA functions in *trans* to promote the expression of myogenin, a key member of the myogenic transcription factor family. More recently, Mueller et al. [32] identified a lncRNA transcribed upstream of *MyoD* named MUNC (MyoD upstream non-coding RNA), and demonstrated that one of the spliced isoforms of MUNC was ^{DRR}RNA. Consistent with the results of Mousavi et al. [59], experimentally decreasing MUNC expression blocked myoblast differentiation, further highlighting the role of enhancer-associated lncRNAs during myogenesis [32]. These results suggest that lncRNAs exhibit regulatory function through *cis*-acting or *trans*-acting mechanisms in skeletal muscle biology and diseases.

All the studies mentioned above have demonstrated that lncRNAs are an integral part of the regulatory network of muscle biology. The present study provides evidence for the role of lncRNAs in skeletal muscle development in goat, which is a starting point for understanding the regulatory mechanisms in which they are involved. The identification of the lncRNAs has greatly improved the annotation of the goat reference genome. Further, we believe that the putative lncRNAs may contribute to a better understanding of the biological basis of regulatory interactions amongst mRNA, miRNA, and lncRNA.

Conclusions

We elucidated skeletal muscle lncRNA profiles of fetal and postnatal goats by RNA-seq analysis and identified and characterized caprine lncRNAs that may be involved in skeletal muscle development in goat. This study provides a catalog of goat muscle lncRNAs that will help in understanding their regulatory roles in goat muscle development. In future studies, we plan to investigate the functions of some of these lncRNAs to provide basic information that will add to the understanding of the regulatory mechanisms associated with goat muscle development at the molecular level.

Methods

Ethics statement

The methods used in this study were performed in accordance with the guidelines of Good Experimental Practices adopted by the Institute of Animal Science (Sichuan Agricultural University, Chengdu, China). All surgical procedures involving goats were performed according to the approved protocols of the Biological Studies Animal Care and Use Committee, Sichuan Province, China.

Animal and tissue preparation

Jianzhou big-eared goats were used in this study. Three pregnant ewes at each developmental stage were subjected to caesarean section to collect the fetuses at 45, 60, 105 days of gestation, and three female lambs were collected at the third day after birth. *Longissimus dorsi* muscle samples were collected at these four developmental stages: three gestation stages (E45, E60, and E105) and one postnatal stage (B3). Three biological replicates for E45, E105, and B3, and two biological replicates for E60 were collected. The eleven samples were immediately frozen in liquid nitrogen for RNA extraction.

RNA extraction, library construction, and sequencing

Total RNA was isolated from the 11 libraries using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. RNA degradation and contamination were monitored on 1 % agarose gels. RNA purity was checked using a NanoPhotometer® spectrophotometer (IMPLEN, Los Angeles, CA, USA). RNA concentration was measured using a Qubit® RNA Assay Kit in a Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). RNA integrity was assessed using a RNA Nano 6000 Assay Kit in a Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Only samples that had RNA Integrity Number (RIN) scores > 8 were used for sequencing. A total of 3 µg RNA per sample was used as input material for RNA sample preparation.

First, rRNA was removed using an Epicentre Ribo-zero rRNA Removal Kit (Epicentre, Madison, WI, USA), and

the rRNA-free residue was obtained by ethanol precipitation. Subsequently, high strand-specificity libraries were generated using the rRNA-depleted RNA and a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA) following the manufacturer's recommendations. Briefly, the rRNA-depleted RNA was fragmented using divalent cations under elevated temperature in NEBNext. First-strand cDNA was synthesized using random hexamer primers and M-MuLV reverse transcriptase (RNase H⁻). Subsequently, second-strand cDNA synthesis was performed using second-strand synthesis reaction buffer, DNA polymerase I, and RNase H. Remaining overhangs were converted into blunt ends by exonuclease/polymerase activity. After adenylation of the 3' ends of the DNA fragments, NEBNext adaptors with hairpin loop structures were ligated to the fragments to prepare them for hybridization. To select cDNA fragments that are 150–200 bp in length, the fragments in each of the library were purified with an AMPure XP system (Beckman Coulter, Brea, CA, USA). Then 3 µl USER Enzyme (NEB, Ipswich, MA, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. The qPCRs were performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. The PCR products were purified (AMPure XP system) and library quality was assessed on an Agilent Bioanalyzer 2100 system. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using a TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. After cluster generation, the libraries were sequenced on an Illumina HiSeq 2500 platform and 125-bp long paired-end reads were generated.

Transcriptome assembly

Clean data were obtained by removing reads containing adapters, reads containing over 10 % of poly(N), and low-quality reads (>50 % of the bases had Phred quality scores ≤ 10) from the raw data. The Phred score (Q20) and GC content of the clean data were calculated. All the downstream analyses were based on the high quality clean data. Goat reference genome and gene model annotation files were downloaded from NCBI database (CHIR_1.0, NCBI) [60]. Index of the reference genome was built using Bowtie v2.0.6 [61, 62] and paired-end clean reads were aligned to the reference genome using TopHat v2.0.14 [63]. The mapped reads from each library were assembled with Cufflinks v2.2.1 [64]. Cufflinks was run with 'min-frags-per-transfrag = 0' and '-library-type fr-firststrand', and other parameters were set as default.

Filtering pipeline to identify multiple exon lncRNAs

We filtered the assembled novel transcripts from the different libraries to obtain putative lncRNAs following

the steps in the pipeline (Fig. 1) as follows. (1) Single exon transcripts and transcripts < 200-bp long were removed. (2) The remaining transcripts that overlapped (>1 bp) with goat protein-coding genes were removed. (3) Transcripts that were likely to be assembly artifacts or PCR run-on fragments according to the class code annotated by cuffcompare [65] were removed. Among the cuffcompare classes, only transcripts annotated as “i”, “u”, and “x” representing novel intronic, intergenic, and antisense transcripts respectively, were retained. (4) The Coding Potential Calculator (CPC) [66], Coding-Non-Coding-Index (CNCI) [67], and phylogenetic codon substitution frequency (PhyloCSF) [68] tools were used to assess the coding potential of the remaining transcripts, and transcripts with CPC score > 0, CNCI score > 0, and PhyloCSF Max_score > 100 were removed. (5) The remaining transcripts that contained a known protein-coding domain were removed. To accomplish this, each transcript sequence was translated in all six reading frames, then HMMER was used to exclude transcripts with a translated protein sequence that had a significant hit in the Pfam (PfamA and PfamB) database, release 28.0 [69]. (6) The remaining transcripts that belonged to known classes of small RNAs (including snRNA, tRNAs, and miRNAs) were removed using Rfam release 12.0 [70].

Expression analysis

Cuffdiff 2.1.1 [71] was used to calculate FPKM scores for the lncRNAs and coding genes in each library. Differentially expressed lncRNAs between any two libraries were identified by edgeR (release 3.2) [72]. We used a P -adjust < 0.01 and an absolute value of the $|\log_2(\text{fold change})| > 2$ as the threshold to evaluate statistical significance of lncRNA expression differences. We analyzed the clusters obtained by systematic analysis for all differentially expressed lncRNAs in eleven libraries using the Heatmaps software package in R [73].

Target gene prediction and functional enrichment analysis

Cis-acting lncRNAs target neighboring genes [74, 75]. We searched for coding genes 10-kb upstream and downstream of all the identified lncRNAs and then predicted their functional roles as follows. The names of the neighbor genes were used to form a gene list that was input into DAVID software [46] for GO analysis. A KEGG enrichment analysis of the predicted target genes was performed with KOBAS software [47] using a hypergeometric test. GO terms and KEGG pathways with $P < 0.05$ were considered significantly enriched.

Co-expression analysis

We used the expression levels of the identified lncRNAs and the known protein-coding genes from the four

different developmental stages to analyze the co-expression of lncRNAs and protein-coding genes. We calculated Pearson's correlation coefficients between the expression levels of 1747 lncRNAs and 19,846 protein-coding transcripts with custom scripts ($r > 0.95$ or $r < -0.95$). Then, we performed a functional enrichment analysis of the candidate lncRNA target genes using DAVID and KOBAS software. Significance was calculated using the Expression Analysis Systematic Explorer (EASE) test method (P -value < 0.05 was considered significant).

Validation of differentially expressed lncRNAs by qPCR

Primers for the lncRNAs and internal control genes (Additional file 10) were designed using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). The expression levels of the lncRNAs were normalized to *ACTB*, *YWHAZ*, and *HPRT1*. Total RNA was converted to cDNA using a PrimeScript™ RT Reagent Kit with gDNA Eraser (TAKARA, Dalian, China), with oligo(dT) and random hexamer primers included in the kit. The qPCR was performed according to the SYBR Premix Ex Taq™ II instructions (TAKARA). The reaction volume contained 10 μl SYBR Premix Ex Taq™ II, 0.8 μl of 10 μM forward and reverse primers, 1.6 μl template cDNA, and dH₂O to a final volume of 20 μl. The reactions were performed on a CFX96 Real-Time PCR System (Bio-Rad, CA, USA) as follows: 95 °C for 2 min, followed by 39 cycles of 95 °C for 10 s, and 10 s at the T_m indicated in Additional file 10: Table S8. Melting curve analysis was performed from 65 to 95 °C with increments of 0.5 °C. Amplifications were performed in triplicate for each sample. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [76], and data were expressed as least square mean \pm standard error of the mean (SEM).

Statistical analyses

All data were analyzed using one-way analysis of variance (ANOVA) to test homogeneity of variances via Levene's test followed by Student's *t*-test analyses in SAS software version 9.0 (SAS, Cary, North Carolina, USA). The significance level for the statistical analysis was $P < 0.05$.

Additional files

Additional file 1: Table S1. Summary of RNA-seq data and reads mapped to the *Capra hircus* reference genome. (XLSX 13 kb)

Additional file 2: Table S2. List of lncRNAs identified in the goat skeletal muscle libraries. (XLSX 171 kb)

Additional file 3: Figure S1. Chromosome distribution of lncRNAs identified in goat skeletal muscle. (TIF 120 kb)

Additional file 4: Table S3. List of differentially expressed lncRNAs from four skeletal muscle developmental stages. (XLSX 122 kb)

Additional file 5: Figure S2. Hierarchical clustering analysis of all expressed lncRNAs from 11 libraries. (TIF 124 kb)

Additional file 6: Table S4. Protein-coding genes detected 10-kb upstream and downstream of the differentially expressed lncRNAs. (XLSX 149 kb)

Additional file 7: Table S5. Functional enrichment analysis of protein-coding genes targeted by *cis*-acting lncRNAs. (XLSX 47 kb)

Additional file 8: Table S6. Co-expression analysis between protein-coding genes and lncRNAs (Pearson's correlation coefficients > 0.95 and < -0.95). (XLSX 16453 kb)

Additional file 9: Table S7. Functional enrichment analysis of protein-coding genes co-expressed with lncRNAs. (XLSX 160 kb)

Additional file 10: Table S8. Primers used in the qPCR analysis. (XLSX 9 kb)

Additional file 11: The transcript annotation information for the lncRNAs. (GTF 2080 kb)

Acknowledgements

We thank all contributors of the present study. We also thank Mr. Cheng Zufu and his colleagues at Gene Denovo Ltd., Co (Guangzhou, China) for assistance in data processing.

Funding

This work was supported by Sichuan Province Science and Technology Support Program (2014NZ0077, 2015NZ0112 and 16ZC2849).

Availability of data and materials

The raw data we obtained from RNA-seq are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (accession number: SRR3567144). The transcript annotation information for the lncRNAs have been provided in form of a gtf file (Additional file 11).

Authors' contributions

SYZ conceived the research, performed data analysis, and drafted the manuscript. YD conducted the animal experiments and sampling. WZ conducted the qPCR validation. LL and JZG performed part of the data analysis. TZ and LJW participated in the experimental design and surgical processes. HPZ provided the experimental environment and edited the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 31 May 2016 Accepted: 10 August 2016

Published online: 22 August 2016

References

- He C, Hu H, Wilson KD, Wu H, Feng J, Xia S, et al. Systematic characterization of long noncoding RNAs reveals the contrasting coordination of *Cis*- and *trans*-molecular regulation in human fetal and adult hearts. *Circ Cardiovasc Genet*. 2016;9(2):110–8.
- Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*. 2015;47(3):199–208.
- Lv J, Liu H, Yu S, Liu H, Cui W, Gao Y, et al. Identification of 4438 novel lincRNAs involved in mouse pre-implantation embryonic development. *Mol Genet Genomics*. 2015;290(2):685–97.
- Goff LA, Groff AF, Sauvageau M, Traves-Gibson Z, Sanchez-Gomez DB, Morse M, et al. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*. 2015;112(22):6855–62.
- Zhu JG, Shen YH, Liu HL, Liu M, Shen YQ, Kong XQ, et al. Long noncoding RNAs expression profile of the developing mouse heart. *J Cell Biochem*. 2014;115(5):910–8.
- Lv J, Huang Z, Liu H, Liu H, Cui W, Li B, et al. Identification and characterization of long intergenic non-coding RNAs related to mouse liver development. *Mol Genet Genomics*. 2014;289(6):1225–35.
- Liang M, Li W, Tian H, Hu T, Wang L, Lin Y, et al. Sequential expression of long noncoding RNA as mRNA gene expression in specific stages of mouse spermatogenesis. *Sci Rep*. 2014;4:5966.
- Koufariotis LT, Chen YP, Chamberlain A, Vander Jagt C, Hayes BJ. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS One*. 2015;10(10):e0141225.
- Billerey C, Boussaha M, Esquerre D, Rebours E, Djari A, Meersseman C, et al. Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 2014;15:499.
- Wang Y, Xue S, Liu X, Liu H, Hu T, Qiu X, et al. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep*. 2016;6:20238.
- Ran M, Chen B, Li Z, Wu M, Liu X, He C, et al. Systematic identification of long noncoding RNAs in immature and mature porcine testes. *Biol Reprod*. 2016;94(4):77.
- Zhao Y, Li J, Liu H, Xi Y, Xue M, Liu W, et al. Dynamic transcriptome profiles of skeletal muscle tissue across 11 developmental stages for both Tongcheng and Yorkshire pigs. *BMC Genomics*. 2015;16:377.
- Zhao W, Mu Y, Ma L, Wang C, Tang Z, Yang S, et al. Systematic identification and characterization of long intergenic non-coding RNAs in fetal porcine skeletal muscle development. *Sci Rep*. 2015;5:8957.
- Bakhtiarzadeh MR, Hosseinpour B, Arefnezhad B, Shamabadi N, Salami SA. In silico prediction of long intergenic non-coding RNAs in sheep. *Genome*. 2016;59(4):263–75.
- Chen YA, Aravin AA. Non-Coding RNAs in transcriptional regulation. *Curr Mol Biol Rep*. 2015;1(1):10–8.
- Vance KW, Ponting CP. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet*. 2014;30(8):348–55.
- Bonasio R, Shiekhattar R. Regulation of transcription by long noncoding RNAs. *Annu Rev Genet*. 2014;48:433–55.
- Morlando M, Ballarino M, Fatica A, Bozzoni I. The role of long noncoding RNAs in the epigenetic control of gene expression. *ChemMedChem*. 2014;9(3):505–10.
- Cao J. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online*. 2014;16:11.
- Mercer TR, Mattick JS. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol*. 2013;20(3):300–7.
- Grote P, Herrmann BG. Long noncoding RNAs in organogenesis: making the difference. *Trends Genet*. 2015;31(6):329–35.
- Mathieu EL, Belhocine M, Dao LT, Puthier D, Spicuglia S. Functions of lncRNA in development and diseases. *Med Sci (Paris)*. 2014;30(8–9):790–6.
- Dey BK, Mueller AC, Dutta A. Long non-coding RNAs as emerging regulators of differentiation, development, and disease. *Transcription*. 2014;5(4):e944014.
- Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–69.
- Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7–21.
- Dey BK, Pfeifer K, Dutta A. The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev*. 2014;28(5):491–501.
- Tang JY, Lee JC, Chang YT, Hou MF, Huang HW, Liaw CC, et al. Long noncoding RNAs-related diseases, cancers, and drugs. *Sci World J*. 2013;2013:1–7.
- Ling H, Vincent K, Pichler M, Fodde R, Berindan-Neaogoe I, Slack FJ, et al. Junk DNA and the long non-coding RNA twist in cancer genetics. *Oncogene*. 2015;34(39):5003–11.
- Liu XH, Sun M, Nie FQ, Ge YB, Zhang EB, Yin DD, et al. lncRNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer*. 2014;13(1):92.
- Buckingham M. Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr Opin Genet Dev*. 2006;16(5):525–32.
- Wang L, Zhao Y, Bao X, Zhu X, Kwok YK, Sun K, et al. lncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res*. 2015;25(3):335–50.

32. Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, et al. MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis. *Mol Cell Biol*. 2015;35(3):498–513.
33. Han X, Yang F, Cao H, Liang Z. Malat1 regulates serum response factor through miR-133 as a competing endogenous RNA in myogenesis. *FASEB J*. 2015;29(7):3054–64.
34. Ballarino M, Cazzella V, D'Andrea D, Grassi L, Bisceglie L, Cipriano A, et al. Novel long noncoding RNAs (lncRNAs) in myogenesis: a miR-31 overlapping lncRNA transcript controls myoblast differentiation. *Mol Cell Biol*. 2015;35(4):728–36.
35. Legnini I, Morlando M, Mangiacavalli A, Fatica A, Bozzoni I. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell*. 2014;53(3):506–14.
36. Braun T, Gautel M. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nat Rev Mol Cell Biol*. 2011;12(6):349–61.
37. Buckingham M, Rigby PW. Gene regulatory networks and transcriptional mechanisms that control myogenesis. *Dev Cell*. 2014;28(3):225–38.
38. Eng D, Ma H-Y, Gross MK, Kioussi C. Gene networks during skeletal myogenesis. *ISRN Dev Biol*. 2013;2013:1–8.
39. Moncaut N, Rigby PW, Carvajal JJ. Dial M(RF) for myogenesis. *FEBS J*. 2013; 280(17):3980–90.
40. Berkes CA, Tapscott SJ. MyoD and the transcriptional control of myogenesis. *Semin Cell Dev Biol*. 2005;16(4–5):585–95.
41. Naya FJ, Olson E. MEF2: a transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation. *Curr Opin Cell Biol*. 1999;11(6):683–8.
42. Estrella NL, Desjardins CA, Nocco SE, Clark AL, Maksimenko Y, Naya FJ. MEF2 transcription factors regulate distinct gene programs in mammalian skeletal muscle differentiation. *J Biol Chem*. 2015;290(2):1256–68.
43. Buckingham M, Relaix F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu Rev Cell Dev Biol*. 2007;23:645–73.
44. Wang Y, Zhang C, Fang X, Zhao Y, Chen X, Sun J, et al. Identification and profiling of microRNAs and their target genes from developing caprine skeletal muscle. *PLoS One*. 2014;9(5):e96857.
45. Wang YH, Zhang CL, Plath M, Fang XT, Lan XY, Zhou Y, et al. Global transcriptional profiling of longissimus thoracis muscle tissue in fetal and juvenile domestic goat using RNA sequencing. *Anim Genet*. 2015;46(6):655–65.
46. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
47. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. 2011;39(Web Server issue):W316–322.
48. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–27.
49. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–91.
50. Xu TS, Zhang XH, Gu LH, Zhou HL, Rong G, Sun WP. Identification and characterization of genes related to the development of skeletal muscle in the Hainan black goat. *Biosci Biotechnol Biochem*. 2012;76(2):238–44.
51. Tripathi AK, Patel AK, Shah RK, Patel AB, Shah TM, Bhatt VD, et al. Transcriptomic dissection of myogenic differentiation signature in caprine by RNA-Seq. *Mech Dev*. 2014;132:79–92.
52. Ren H, Wang G, Chen L, Jiang J, Liu L, Li N, et al. Genome-wide analysis of long non-coding RNAs at early stage of skin pigmentation in goats (*Capra hircus*). *BMC Genomics*. 2016;17(1):67.
53. Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol*. 2015;11(3):151–60.
54. Ounzain S, Micheletti R, Beckmann T, Schroen B, Alexanian M, Pezzuto I, et al. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. *Eur Heart J*. 2015;36(6):353–368a.
55. Bismuth K, Relaix F. Genetic regulation of skeletal muscle development. *Exp Cell Res*. 2010;316(18):3081–6.
56. Gong C, Li Z, Ramanujan K, Clay I, Zhang Y, Lemire-Brachat S, et al. A long non-coding RNA, lncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev Cell*. 2015;34(2):181–91.
57. Lu L, Sun K, Chen X, Zhao Y, Wang L, Zhou L, et al. Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J*. 2013;32(19):2575–88.
58. Nie M, Deng ZL, Liu J, Wang DZ. Noncoding RNAs, emerging regulators of skeletal muscle development and diseases. *Biomed Res Int*. 2015;2015:1–17.
59. Mousavi K, Zare H, Dell'Orso S, Gronvold L, Gutierrez-Cruz G, Derfoul A, et al. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell*. 2013;51(5):606–17.
60. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol*. 2013;31(2):135–41.
61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
62. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
63. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.
64. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
65. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 2011;27(17):2325–9.
66. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server issue):W345–349.
67. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e166.
68. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27(13):i275–282.
69. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–285.
70. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*. 2015; 43(Database issue):D130–137.
71. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc*. 2012;7(3):562–78.
72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
73. Ploner A. Heatplus: heatmaps with row and/or column covariates and colored clusters. R package version 2.18.0. 2015.
74. Ponjavic J, Oliver PL, Lunter G, Ponting CP. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet*. 2009;5(8):e1000617.
75. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 2010;143(1):46–58.
76. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods*. 2001;25(4):402–8.