ANIMAL BIOSCIENCE

Identification and functional prediction of long non-coding RNAs related to skeletal muscle development in Duroc pigs

Lixia Ma[1], Ming Qin[1], Yulun Zhang[2], Hui Xue[2], Shiyin Li[1], Wei Chen[1], and Yongqing Zeng[1,*]

* Corresponding Author: Yongqing Zeng
Tel: +86-0538-8249222-8311,
Fax: +86-0538-8249222-8311,
E-mail: yqzeng@sdau.edu.cn

[1] Shandong Provincial Key Laboratory of Animal Biotechnology and Disease Control and Prevention, College of Animal Science and Technology, Shandong Agricultural University, Tai'an, Shandong 271018, China
[2] Shandong Ding Tai Animal Husbandry Co. Ltd., Jinan, Shandong 250300, China

ORCID
Lixia Ma
https://orcid.org/0000-0001-9475-7650
Ming Qin
https://orcid.org/0000-0003-2212-2626
Yulun Zhang
https://orcid.org/0000-0003-4229-8207
Hui Xue
https://orcid.org/0000-0002-0867-7126
Shiyin Li
https://orcid.org/0000-0002-3451-1912
Wei Chen
https://orcid.org/0000-0002-2901-5156
Yongqing Zeng
https://orcid.org/0000-0002-0740-2470

Objective: The growth of pigs involves multiple regulatory mechanisms, and modern molecular breeding techniques can be used to understand the skeletal muscle growth and development to promote the selection process of pigs. This study aims to explore candidate lncRNAs and mRNAs related to skeletal muscle growth and development among Duroc pigs with different average daily gain (ADG).

Methods: A total of 8 pigs were selected and divided into two groups: H group (high-ADG) and L group (low-ADG). And followed by whole transcriptome sequencing to identify differentially expressed (DE) lncRNAs and mRNAs.

Results: In RNA-seq, 703 DE mRNAs (263 up-regulated and 440 down-regulated) and 74 DE lncRNAs (45 up-regulated and 29 down-regulated) were identified. In addition, 1,418 Transcription factors (TFs) were found. Compared with mRNAs, lncRNAs had fewer exons, shorter transcript length and open reading frame length. DE mRNAs and DE lncRNAs can form 417 lncRNA-mRNA pairs (antisense, cis and trans). DE mRNAs and target genes of lncRNAs were enriched in cellular processes, biological regulation, and regulation of biological processes. In addition, quantitative trait locus (QTL) analysis was used to detect the functions of DE mRNAs and lncRNAs, the most of DE mRNAs and target genes of lncRNAs were enriched in QTLs related to growth traits and skeletal muscle development. In single-nucleotide polymorphism/insertion-deletion (SNP/INDEL) analysis, 1,081,182 SNP and 131,721 INDEL were found, and transition was more than transversion. Over 60% of percentage were skipped exon events among alternative splicing events.

Conclusion: The results showed that different ADG among Duroc pigs with the same diet maybe due to the DE mRNAs and DE lncRNAs related to skeletal muscle growth and development.

Keywords: Duroc Pigs; Long Non-coding RNAs; mRNAs; Skeletal Muscle

INTRODUCTION

Skeletal muscle occupies 45% to 60% of the animal body weight, it is the most abundant tissue in animals, and it is also one of the most important production traits for the growth and development of livestock and poultry animals. In medicine, abnormal regulation of skeletal muscle can lead many different types of muscle diseases, such as, myosarcoma, muscular atrophy, dystrophy, and muscular hypertrophy [1]. Failure of skeletal muscle development before birth can lead to embryonic death, and failure to repair or maintain skeletal muscle after birth can lead to a decline in the quality of life, and sometimes even death [2]. Therefore, exploring the mechanisms of skeletal muscle growth and development will help improve livestock production performance. High-throughput sequencing technology, genome-wide association study (WGAS), whole genome bisulfite sequencing (WGBS) and RNA-sequencing (RNA-seq) are gradually being applied to the study of skeletal muscle [3-5], and a large number of key candidate genes related to the growth

and development of skeletal muscle have been identified [6,7].

The length of lncRNA is more than 200 nt, and the most of lncRNAs are transcribed by polymerase II, modified by splicing, capping, and tailing, and have characteristics similar to mRNA, including 5'cap and 3'polyA tail structure [8]. But compared with mRNA, it has poor conservation, low expression abundance and tissue specificity [9]. In recent years, with the continuous deepening of lncRNA function research, more and more studies have proved that lncRNAs can regulate muscle growth and differentiation, cell cycle and cell apoptosis [10,11]. lncRNA-Six1 exerts cis-regulation on Six1 gene encoding protein, and encodes micropeptide to active Six1 genes, thereby promoting cell proliferation and participating in muscle growth [12]. lncRNA-Dum is in skeletal myoblast cells, its expression is dynamically regulated during myogenesis. It is also transcriptionally introduced by MyoD binding during myoblast differentiation. The results show that it can promote myoblast differentiation and damage-induced muscle regeneration [13]. lnc133b can regulate bovine skeletal muscle satellite cell proliferation and cell apoptosis by mediating "sponge" miR-133b [14]. But RNA-seq is rarely applied to studies of full or half-siblings, this study may provide a novel method and clarify the relative precise mechanisms.

Duroc pigs are famous for their high growth rate, feed conversion efficiency, and lean meat percentage [15,16]. Duroc, as a typical lean pig breed, has good growth and meat production performance before 110 kg body weight, but its growth and meat production performance gradually decreases after 110 kg body weight. Therefore, the selection of 110 to 130 kg body weight can continue to maintain good growth and meat production performance, that is, cultivating a new breed of high-yield and high-quality meat pigs with large body size, which would have a great effect on improving the economic benefits of the pig industry. Therefore, this study aims to explore candidate lncRNAs and mRNAs related to skeletal muscle growth and development among Duroc pigs with different average daily gain (ADG).

MATERIALS AND METHODS

Ethics statement
All animal care and treatment procedures were conducted in strict accordance with the Animal Ethics Committee of Shandong Agricultural University, China, and performed in accordance with the Committee's guidelines and regulations (Approval No.: 2004006).

Animals
Duroc pigs came from a core breeding farm, with the measurement data in the pig herd to 30 to 110 kg body weight (individuals in the top 30% of ADG), and the performance

measurement was continued to about 130 kg body weight. According to the ADG, 8 pigs were selected and divided into two groups: 4 high-ADG group (774.89 g) (H group) and 4 low-ADG group (658.77 g) (L group), and each pair of high and low groups were half siblings. The longissimus dorsi muscle (LDM) tissues were sampled and snap-frozen in liquid nitrogen for extraction of total RNA.

RNA extraction, strand-specific library construction and sequencing
Total RNA was extracted from LDM tissues with Trizol reagent kit (Invitrogen, Carlsbad, CA, USA) according to the instructions. RNA quality (Supplementary Figure S5) was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and checked with RNase free agarose gel electrophoresis. The enriched mRNAs and ncRNAs were fragmented into short fragments with fragmentation buffer after ribosome RNA (rRNA) removed. First-strand was transcribed with random primers. Second-strand of cDNA was synthesized with DNA polymerase I, RNase H, dUTP and buffer. Then, the cDNA was purified with QiaQuick polymerase chain reaction (PCR) extraction kit (Qiagen, Venlo, The Netherlands), end repaired, poly(A) added, and ligated to Illumina sequencing adapters. Then the second-strand cDNA was digested with uracil-N-glycosylase. The digested products were size selected with agarose gel electrophoresis, amplified, and sequenced with Illumina HiSeqTM 4000 (In this study, the paired-end sequencing method of Illumina sequencing was used, and each end was the 150 bp reads inserted into the target sequence, which were sequenced.) by Gene Denovo Biotechnology Co. (Guangzhou, China).

Quality control for raw reads (raw datas) and mapping
Raw reads contained adapters or low quality reads were contained. Thus, reads were further filtered by using fastp [17] (https://github.com/OpenGene/fastp) (version 0.18.0) to get high quality clean reads. The quality control standards were as follows: i) removing reads with adapters; ii) removing reads with more than 10% of unknown nucleotides (N); iii) removing low quality reads with more than 50% of low quality (Q-value≤20) bases. Bowtie2 [18] (version 2.2.8) was used for mapping reads to rRNA database, the mapped reads were removed, and the remaining reads were further used in assembly and analysis of transcriptome. The rRNA removed reads of each sample were then mapped to reference genome by HISTA2 [19], respectively. After aligned with reference genome, on average about 84.10% clean reads were mapped to the reference genome (Ensembl-release104), and 77.60% clean reads were unique mapped, and only 6.5% clean reads were multiple mapped (Supplementary Table S1).

Quantification and differentially expressed transcripts

analysis

Transcript abundances were quantified with StringTie software in a reference-based approach. A FPKM (fragment per kilobase of transcript per million mapped reads) value was calculated to quantify. mRNAs and lncRNAs differential expression analyses were performed with DESeq2 software. The genes with the parameter of p-value below 0.05 and absolute fold change ≥1.5 were considered differentially expressed (DE) mRNAs and lncRNAs.

Exon and intron analysis

The information of introns and exons exists in the structural annotation of the reference genome, so when the transcriptome data is compared with the reference genome, the intron and exon information of the compared genes can be directly called. When reads were aligned with introns, they belong to introns, and when they were aligned with exons, they belong to exons.

Gene ontology and Kyoto encyclopedia of genes and genomes enrichment analysis

Gene ontology (GO) is an international standardized gene functional classification system, which includes three parts: cellular component (CC), biological process (BP), and molecular function. The analysis of Pathway helps to further know about the biological functions of genes. Kyoto encyclopedia of genes and genomes (KEGG) is the main public database about Pathway [20].

lncRNA-mRNA association analysis

To reveal the interaction between antisense lncRNAs and mRNAs, the software RNAplex [21] (version 0.2) was used to predict the complementary binding between antisense lncRNAs and mRNAs. The ViennaRNA package is a program in RNAplex software [22].

One of the functions of lncRNAs is cis-regulation of their neighboring genes on the same allele. The up-stream lncRNAs which intersect a promoter or other cis-elements may regulate gene expression in transcriptional or post-transcriptional level. The down-stream or 3'UTR region lncRNAs may have other regulatory functions. Thus, lncRNAs which had been previously annotated as "unknown region" were annotated again. lncRNAs in less than 100 kb up/down stream of a gene were likely to be cis-regulators.

Another function of lncRNAs is trans-regulation of co-expressed genes not adjacent to lncRNAs. We analyzed the correlation of expression between lncRNAs and protein-coding genes to identify target genes of lncRNAs.

Single-nucleotide polymorphism and insertion-deletion analysis

The GATK [23] (version 3.4-46) was used for calling variants of transcripts, and ANNOVAR [24] was used for single-nucleotide polymorphism/insertion-deletion (SNP/InDel) annotation. The following criteria were used to screen reliable editing sites from SNP sites: i) Removing the low quality SNP by GATK. ii) Correcting the SNP around INDEL region. iii) Choosing non-overlapping SNP in EXON and UTR region. iv) Choosing SNP with reference reads≥2 and variant reads ≥3. v) Choosing SNP with the variation frequency between 0.1 and 0.9 [25,26].

Alternative splicing (AS) analysis

rMATS [27] (version 4.0.1) (http://rnaseq-mats.sourceforge.net/index.html) was used to identify alternative splicing (AS) events and analyze differential AS events. There are two methods to analyze AS by rMATS. Junction count (JC) only considered the number of splicing events spanning a splice site. Reads On Target and Junction Count (JCEC) considered not only the reads spanning the splice site, but also the number of splicing events for the reads that do not span the splice site (exon counts). At the same time, to count AS events matched to exonic regions, two methods were used to analyze AS events. AS events with a false discovery rate <0.05 in a comparison as significant AS events.

Quantitative real-time polymerase chain reaction analysis

The total RNA was reverse-transcribed into cDNA by Prime-Script RT reagent kit (TaKaRa, Dalian, China), and the analyzed by SYBR Green Pro Taq HS Premix (Accurate Biotechnology (Hunan) Co., Ltd, Changsha, China). Primers were compounded by Sangon Biotech (Shanghai, China), and sequences of primers are shown in Table 1. β-Actin was used as a housekeeping gene. The fold change in expression was the obtained by $2^{-\Delta\Delta CT}$ method, $\Delta\Delta CT = (CT_{Target\ gene} - CT_{\beta-actin})_{H\ group} - (CT_{Target\ gene} - CT_{\beta-actin})_{L\ group}$.

RESULTS

Identification of DE mRNAs and DE lncRNAs

In this study, 21,353 mRNAs were identified, including 61 novel genes and 21,292 known genes. 1,418 transcription factors (TFs) were found, the largest number is zf-C2H2 family (493 genes), 195 TFs belong to OTX family (Supplementary Figure S1). A total of 703 DE mRNAs were identified, including 263 up-regulated genes and 440 down-regulated genes (Figure 1A, B; Supplementary Table S9). In addition, a total of 10,178 lncRNAs were identified, including 8,478 Intergenic lncRNAs (Supplementary Figure S2A), and 479 lncRNAs were Scaffold, and the rest were located on the known chromosomes of pigs. This result showed that the distribution of the identified lncRNAs on the chromosomes did not have obvious bias (Supplementary Figure S2B). Compared

Table 1. Primers sequences

Gene	Sequence (5'-3')
lncRNA ENSSSCT00000053775-F	TTAGGTACGCCTCCCCAACTTCC
lncRNA ENSSSCT00000053775-R	GGCAACAGATTCCAGACCCAGATC
lncRNA ENSSSCT00000078829-F	CACCTGCCATTGTCCATCTCTGTC
lncRNAENSSSCT00000078829-R	CGGAGCTTCCAGACTAGGAGAGG
lncRNAENSSSCT00000087634-F	GCATTCCACAGACATCAGAGAGTCC
lncRNAENSSSCT00000087634-R	GTAGAAGCAGAGAAGGCAGCAAGG
lncRNAENSSSCT00000074339-F	CACAGCCACAGCCACAACAG
lncRNAENSSSCT00000074339-R	TAAGGATCCAGCGTTGCCGT
lncRNAENSSSCT00000074036-F	TGGAGACGCCCTTGACCAAG
lncRNAENSSSCT00000074036-R	GTGATGGGCTGGCTGGAGAT
MTMR7-F	AGACTCTTCACTCGGGACGGTAAC
MTMR7-R	GGTGGTGCAGGAGGAGTAGGAG
PRKCB-F	ATTCACCGCCCGCTTCTTCAAG
PRKCB-R	AACAGCAGACTTGGCACTGGAATC
ETV7-F	AAGAAGGAACCTGGGCAGAAACTC
ETV7-R	GTCTGTCCTGCTCCTGGCTCTC
TNC-F	CTGCCACCGAATACACGCTGAG
TNC-R	GCTGTTTCCGACTGAACCTCTGTAG
β-actin-F	CCAGGTCATCACCATCGGCAAC
β-actin-R	CAGCACCGTGTTGGCGTAGAG

MTMR7, myotubularin related protein 7; *PRKCB*, protein kinase C beta; *ETV7*, ETS variant transcription factor 7; *TNC*, tenascin C.

with mRNAs, the average number of lncRNAs exon was 2.88, which was much lower than the number of mRNAs (9 exons) (Supplementary Figure S3A,B). The average length of mRNAs open reading frame (ORF) was 1,762 aa (Supple-
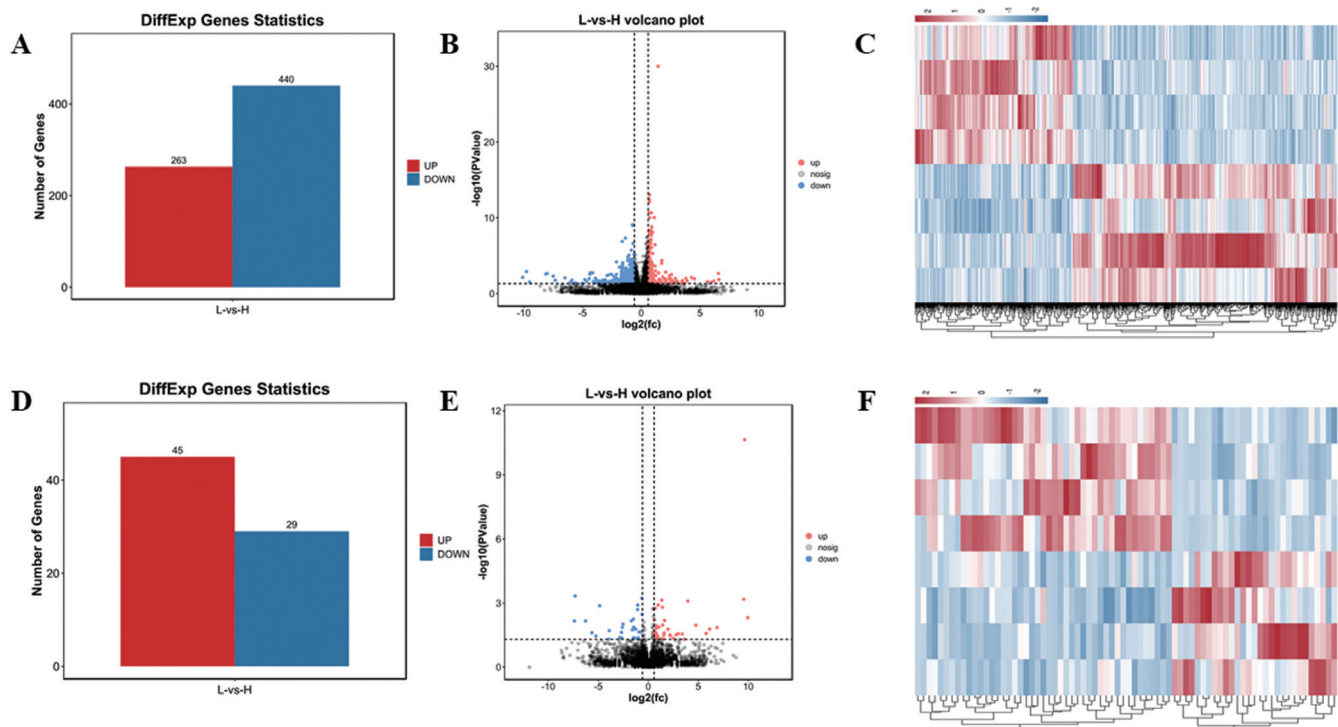
Figure 1. Statistics of DE mRNAs and DE lncRNAs. (A), (D) Statistics of 703 DE mRNAs and 74 DE lncRNAs. Red represents up-regulation, blue represents down-regulation. (B), (E) Volcano plot analysis of 703 DE mRNAs and 74 DE lncRNAs. Red represents up-regulation, blue represents down-regulation. (C), (F) Heat map analysis of 703 DE mRNAs and 74 DE lncRNAs. Each row represents a sample, and each column represents a DE mRNA or DE lncRNA. Red and blue gradients indicate increase and decrease in gene expression abundance, respectively. DE, differentially expressed.

mentary Figure S3D). The length of lncRNAs (2,747 nt) was shorter than mRNA (3,442 nt) (Supplementary Figure S3E, 3F). In addition, the relative expression level of lncRNAs was lower than mRNA (Supplementary Figure S3G, H). A total of 74 DE lncRNAs were identified, including 45 up-regulated lncRNAs and 29 down-regulated lncRNAs (Figure 1D, E; Supplementary Table S8). Hierarchical clustering showed that the expressions of genes and lncRNAs were distinguishable between H group and L group, indicating a significant difference between H group and L group (Figure 1C, 1F).

To verify RNA-seq results, five DE mRNAs and five lncRNAs were selected to perform by quantitative real-time polymerase chain reaction (qRT-PCR). The results showed that qRT-PCR results agreed with those in RNA-seq (Figure 2). The results indicated that DE mRNAs and DE lncRNAs identified from RNA-seq were reliable.

lncRNA-mRNA association analysis

lncRNA-mRNA association analysis can be divided into three parts: antisense lncRNA analysis, cis regulation lncRNA analysis and trans regulation lncRNA analysis [28].

All lncRNAs/mRNAs and DE lncRNAs/mRNAs were analyzed (Supplementary Table S2). The results included one DE antisense lncRNA and one mRNA (Table 2).

One of the functions of lncRNAs is cis regulation of their neighboring genes on the same allele. As shown in Table 3, two DE lncRNAs and DE two mRNAs were identified (Table 3; Supplementary Table S3).

Another function of lncRNAs is trans regulation of co-expressed genes not adjacent to lncRNAs. Pearson correlation coefficient was used for samples≥6, and protein-coding genes with absolute correlation was more than 0.999. As shown in Table 4, 57 DE lncRNAs and 241 DE mRNAs were related, and 414 pairs lncRNA-mRNA were found.

Gene ontology enrichment analyses of DE mRNAs and target genes of DE lncRNAs

To understand the functions of DE mRNAs and DE lncRNAs, the DE mRNAs and target genes of DE lncRNAs were annotated by GO enrichment analysis. The results showed that DE mRNAs were mainly enriched in BP, including cellular process, metabolic process, biological regulation, and regu-
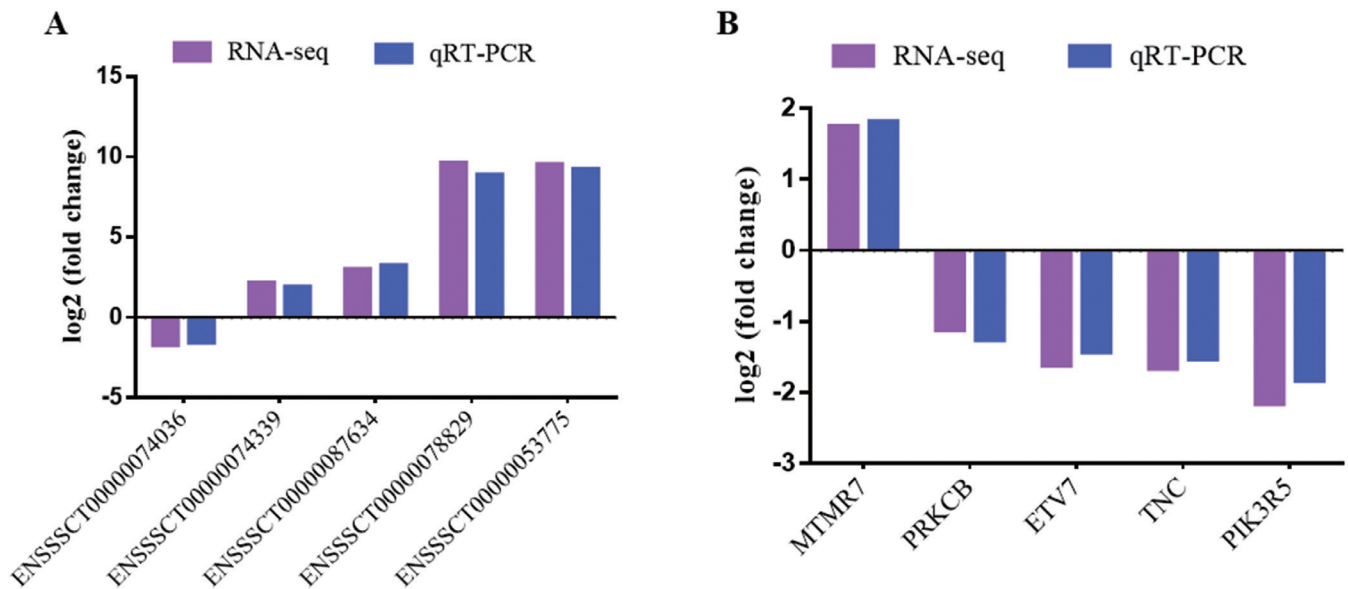

Figure 2. Validation of DE lncRNAs and DE mRNAs. (A) DE lncRNAs. (B) DE mRNAs. Purple represents RNA-seq result, blue represents qRT-PCR result. DE, differentially expressed; qRT-PCR, quantitative real-time polymerase chain reaction.

Table 2. Differentially expressed antisense regulation results

lncRNA_ID	mRNA_ID	Symbol	Description
MSTRG.6467.1	ENSSSCG00000036785	MTMR7	myotubularin related protein 7

Table 3. Differentially expressed cis regulation results

lncRNA_ID	mRNA_ID	Symbol	Description
ENSSSCT00000087573	ENSSSCT00000087573	-	-
MSTRG.12078.1	ENSSSCG00000022230	CD9	CD9 moleculeprotein 7

Table 4. Trans regulation results

List	lncRNA	mRNA	Pair
Diff	57	241	414

lation of BP. Cellular component was mainly including cell, cell part, organelle, and membrane part. Molecular function was mainly including binding, catalytic and molecular transducer activity (Figure 3A). The first 20 GO terms were mainly involved in biological process (Supplementary Figure S4A) (Q≤0.05). Target genes of lncRNAs (antisense, cis and trans regulation) were mainly enriched in cellular process, cell, cell parts and binding (Figure 3B, C, D). The first 20 GO terms were shown in Supplementary Figure S4B, S4C, S4D (Q≤0.05).

KEGG analysis of DE mRNAs and target genes of DE lncRNAs

A total of 703 DE mRNAs were enriched in 286 signaling pathways. Complement and coagulation cascades and viral protein interaction with cytokine and cytokine receptor signaling pathways were the two significantly pathways. These DE mRNAs were also significantly enriched in pathways in-cluding chemokine signaling pathway, mitogen-activated protein kinase (MAPK) cascade (p = 0.038), extracellular signal-regulated kinase 1 (ERK1) and ERK2 cascade (p = 0.003) and positive regulation of ERK1 and ERK2 cascade (p = 0.015) (Figure 4A). Target genes of antisense regulation lncRNAs were mainly enriched in pathways including ribosome, huntington disease, thermogenesis, and human cytomegalovirus infection (Figure 4B). Target genes of cis regulation lncRNAs were mainly enriched in alcohol, cyclic guanosine monophosphate protein kinase G -protein kinase G (cGMP-PKG) signaling pathway, plate activation and RNA transport signaling pathway (Figure 4C). while Target genes of trans regulation lncRNAs were mainly enriched in complement and coagulation cascades, chemokine signaling pahway, phagosome and cytokine-cytokine receptor interaction signaling pathway (Figure 4D).

Expression analysis of growth traits and meat quality-relevant QTLs

Skeletal muscle is closely related to pig growth and meat quality. To further study the potential functions of DE mRNAs and DE lncRNAs. Basic Local Alignment Search Tool (blast) comparison was performed on the quantitative
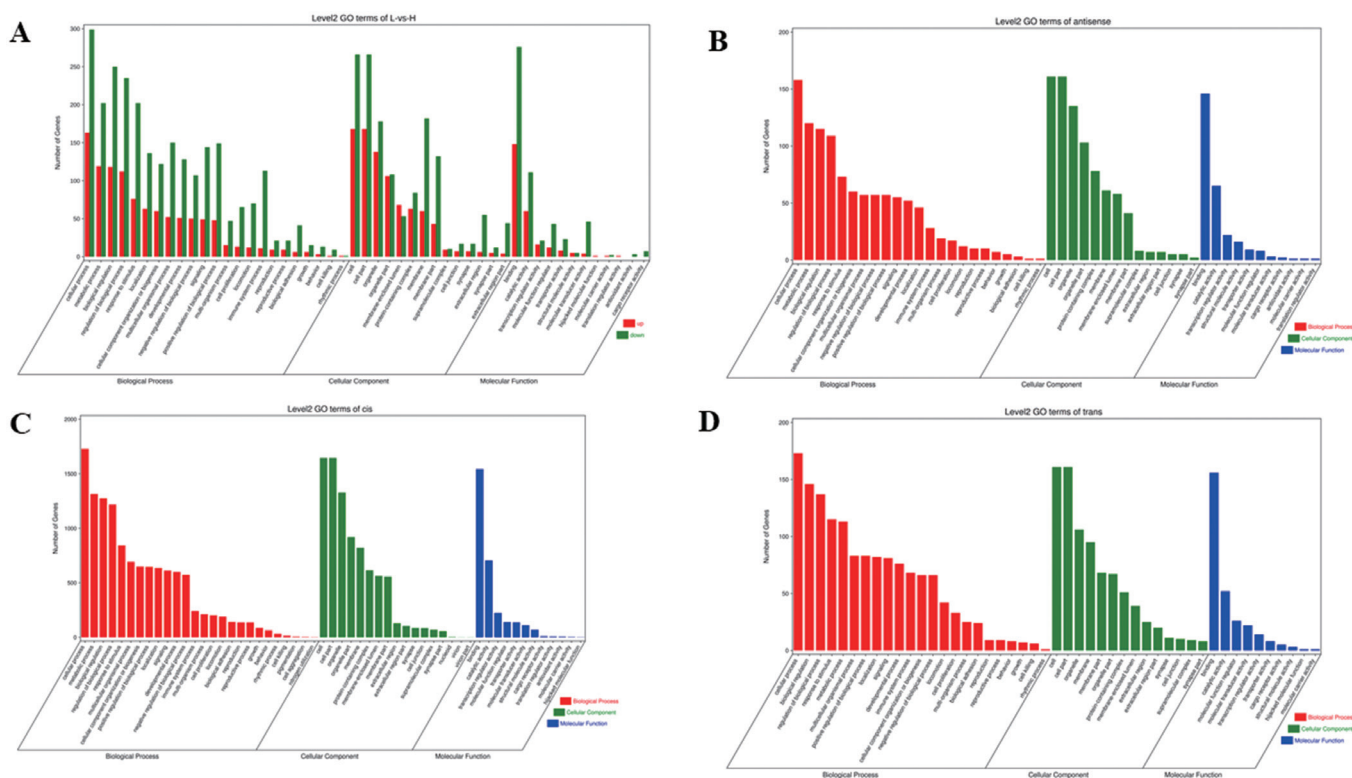
Figure 3. Analysis of GO enrichment. (A) DE mRNAs. The abscissa is the secondary GO term, the ordinate is the number of DE genes in this term, red means up-regulation, green means down-regulation. (B) Target genes of antisense regulation lncRNAs. (C) Target genes of cis regulation lncRNAs. (D) Target genes of trans regulation lncRNAs. (B), (C), (D) The abscissa is the secondary GO term, and the ordinate is the number of genes in this term. GO, gene ontology; DE, differentially expressed. Different colors show different types of GO terms, red represents biological process, green represents cellular component, blue represent molecular function.
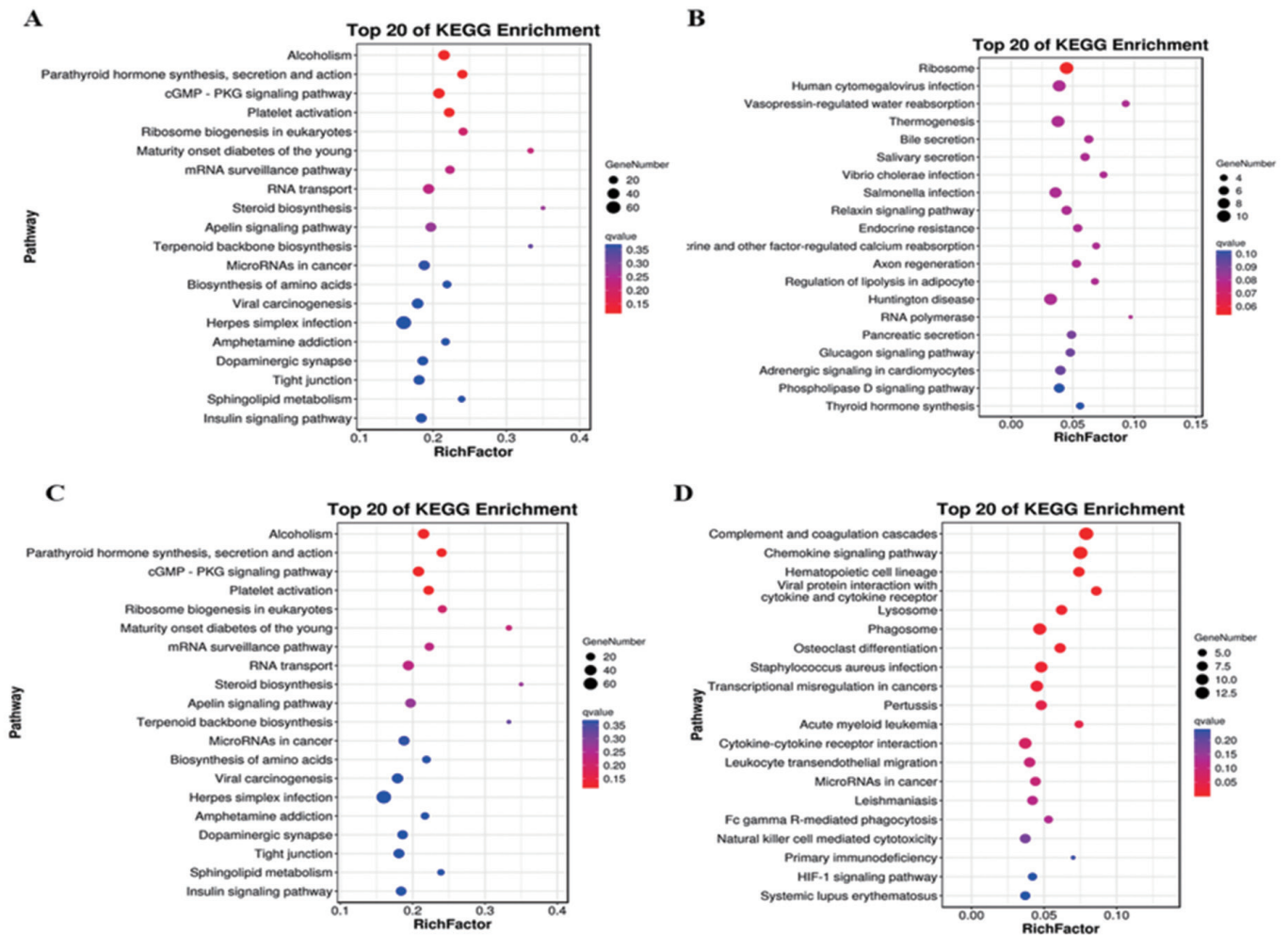
Figure 4. Analysis of KEGG enrichment. (A) DE mRNAs. (B) Target genes of antisense regulation lncRNAs. (C) Target genes of cis regulation lncR-NAs. (D) Target genes of trans regulation lncRNAs. Using the top 20 pathways with the smallest Q value to make a map, the ordinate is the pathway, the abscissa is the enrichment factor, the size of the circle indicates the number of genes, the redder the color, the smaller the Q value. KEGG, Kyoto encyclopedia of genes and genomes; DE, differentially expressed.

trait locus (QTL) (<2 Mb) with high confidence related to pig growth and meat quality traits. If the transcript or QTL interval is more than 50% duplicated, the transcript is located on this QTL. According to the QTL mapping analysis, 652 DE mRNAs were enriched in the QTLs related to growth traits and meat quality, including protein kinase C beta (*PRKCB*) confirmed by qRT-PCR, was enriched in Number of muscle fibers per unit area QTL and Cholesterol level QTL (Table 5). A total of 967 target genes of DE lncRNAs were enriched in QTLs. CD9, as target gene of lncNRA *MSTRG.12078.1*, was enriched in Backfat at first rib QTL and Marbling QTL, and the length of QTL was over 2,321 MB (Table 6).

Single-strand nucleotide polymorphism and insertion-deletion analysis
Variation analysis based on transcriptome sequencing mainly includes SNP and INDEL. In this study, 1,081,182 SNP and 131,721 INDEL were confirmed (Supplementary Table S4).

As shown in Figure 5A, SNPs were mainly located in intronic and intergenic, in addition, SNPs in other locations were also found, such as 3'UTR, exonic, 5'UTR and other locations, but with far less than the intronic and intergenic SNPs. And functions were mainly synonymous single nucleotide variations (SNVs) and nonsynonymous SNVs (Figure 5B). SNP type includes transversion and transition. As shown in Figure 5C, the ratio of transversion was 24.72% (A-to-T 2.49%, T-to-A 2.50%, T-to-G 3.16%, A-to-C 3.19%, G-to-T 3.31%, C-to-A 3.32%, G-to-C 3.36%, C-to-G 3.39%), while transition percentage was 75.28% (C-to-T 17.28%, G-to-A 17.2-%, T-to-C 20.28%, A-to-G 20.43%).

RNA editing is a post transcriptional modification widely existing in eukaryotes. At present, SNVs are common in RNA editing. In this study, the frequency of SNVs was similar in 8 samples (Figure 5D, Supplementary Table S5).

Alternative splicing analysis

Table 5. Differentially expressed mRNAs distribution in chromosome and QTLs regions

Chromosome/ mitochondrion	DE gene number	DE gene number in QTL region	QTL region length (Mb)
1	51	51	272.538
2	75	74	152.534
3	37	36	129.972
4	48	47	126.574
5	38	35	98.508
6	73	67	162.985
7	49	49	126.47
8	27	25	143.374
9	34	33	138.044
10	6	5	67.1919
11	9	9	77.1184
12	31	29	58.0079
13	51	51	208.259
14	46	42	154.596
15	25	24	137.024
16	18	17	66.7432
17	22	19	55.1875
18	21	21	58.1156
X	25	18	85.3869
Y	1	0	0
MT	8	0	0

DE, differentially expressed; QTL, quantitative trait locus.

Table 6. Target genes of DE lncRNAs distribution in chromosome and QTLs regions

Chromosome/ Mitochondrion	DE gene number	DE gene number in QTL region	QTL region length (Mb)
1	95	95	272.538
2	91	89	152.534
3	59	58	129.972
4	64	60	127.273
5	62	56	98.508
6	108	104	162.985
7	68	68	128.937
8	45	40	143.428
9	47	47	138.044
10	21	20	67.1919
11	15	14	78.2374
12	49	45	58.0079
13	79	79	208.259
14	51	46	154.596
15	41	41	137.024
16	24	24	66.7432
17	38	34	55.1875
18	25	25	58.1156
X	29	22	83.7497
Y	4	0	0
MT	2	0	0

DE, differentially expressed; QTL, quantitative trait locus.
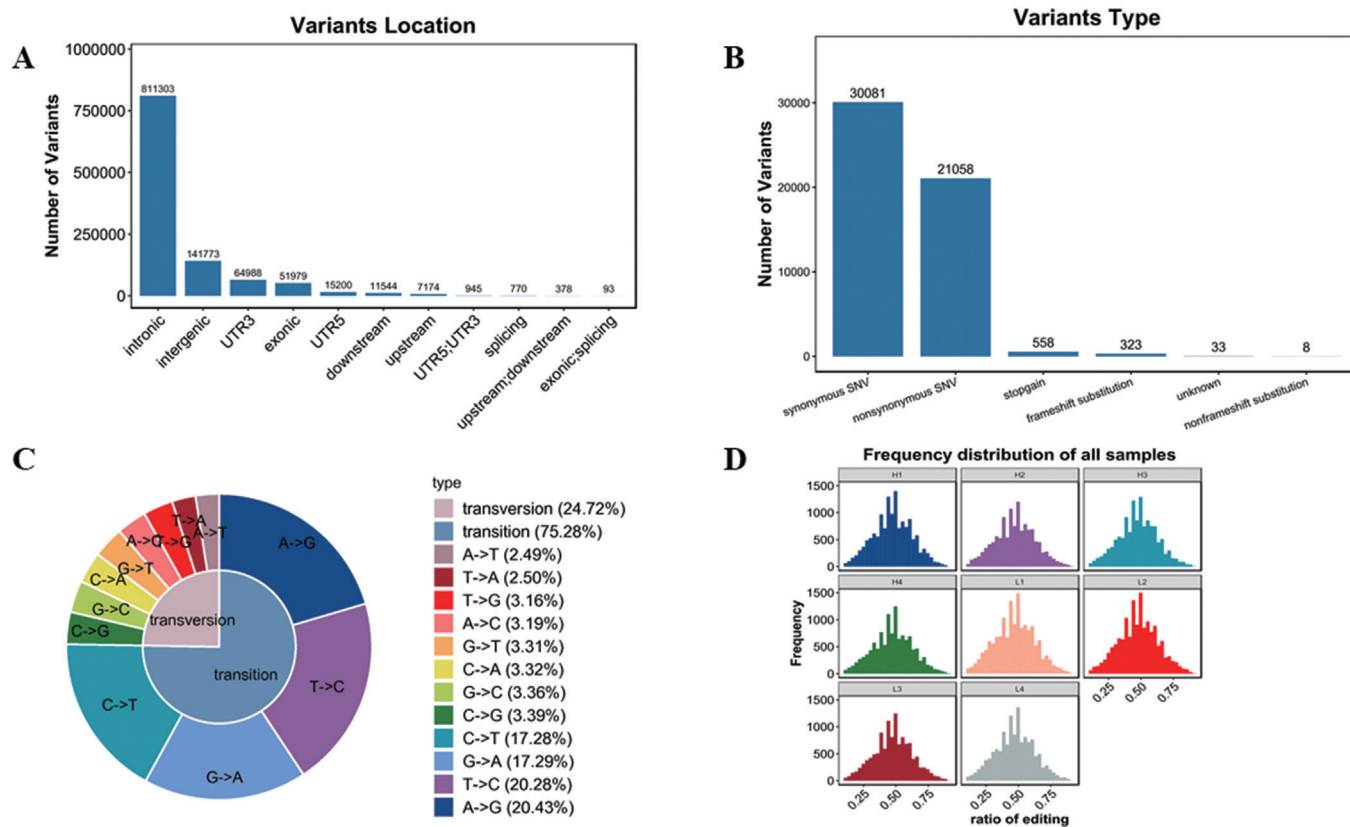
Figure 5. Analysis of variants. (A) Variants location. The abscissa represents variants location, the ordinate represents the number of variants. (B) Variants function. The abscissa represents variants type, the ordinate represents the number of variants. (C) Variants type. (D) Frequency distribution of all samples.

---

to study DE lncRNAs and mechanisms.

In GO and KEGG results, although the top 20 GO terms and pathways have no pathways related to skeletal muscle growth and development, some related pathways were still analyzed, which were identified to be related with skeletal muscle development [40,41]. In this study, it was found that antisense LncRNA MSTRG.6467.1-MTMR7 pair was related to inositol phosphate metabolism and phosphatidylinositol signaling system. The cis LncRNA MSTRG. 12078.1-CD9 were enriched in hematopoietic cell lineage. Although the specific function of the identified DE mRNAs and lncRNAs are unclear, it also provided some signaling pathways for the growth and development mechanism of skeletal muscle.

RNA editing can efficiently change the amino acid sequence of the protein, thereby changing the genetic information the genome template [42]. Abnormal RNA editing is closely related to human diseases such as amyotrophic lateral sclerosis and cancer [43]. The diversity of RNA in eukaryotes mainly comes from A-to-G. In this study, the frequency of transition (75.28%) is significantly higher than transversion (24.72%). And A-to-G was the most frequent, which was similar with previous research [44]. The diversification of RNA and DNA editing types was designed to the development of genetics, molecules, biochemistry, and other aspects to solve biological problems, but there are still many challenges. With the development of high-throughput technology, these problems will gradually be solved.

AS is a crucial factor in increasing the complexity of cell functions [45], the identified AS events will help to better know about the regulation mechanism skeletal muscle development. rMATS [46] was used to analyze AS. SE was the most common of the AS events, which was different from some reports [3,47]. This may be related to the different analysis methods and samples. As shown in a report, myotilin (*MYOT*) lack an entire exon13 [48], In addition, the AS events occurring in *MEF2C* and *MEF2D* have a certain regulation effect on skeletal muscle differentiation and myogenesis [45,49]. The AS events of *MEF2C*, *MEF2D*, and *MYOT* were found in this study, which can provide certain data support for the further study of genes related to skeletal muscle growth and development.

CONCLUSION

In conclusion, 703 DE mRNAs and 74 DE lncRNAs were identified. Compared with mRNAs, lncRNAs had fewer exons, shorter transcript length and ORF length. DE mRNAs and DE lncRNAs can form 417 lncRNA-mRNA pairs (antisense, cis, trans). In addition to cell and cell parts, DE mRNAs and target genes of lncRNAs were also enriched in MAPK cascade and ERK1/ERK2 cascade. In addition, QTL analysis was used to detect the functions of DE mRNAs and lncRNAs,

the most of DE mRNAs and target genes of lncRNAs were enriched in QTLs related to skeletal muscle development and meat quality. In SNP/INDEL analysis, 1,081,182 SNP and 131,721 INDEL were found, and transition was more than transversion. Over 60% of percentage were SE events among AS events. The large amount of data identified by RNA-seq in this study requires more technical verification to select high-quality pigs, and to further provide a theoretical basis for pig genetic improvement and meat quality research.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The data used to support the findings this study is available from the corresponding author upon request.

Supplementary file is available from: https://doi.org/10.5713/ab.22.0020

Supplementary Table S1. Statistics for filtering and mapping reads
Supplementary Table S2. Analysis of antisense regulation lncRNAs
Supplementary Table S3. Analysis of cis regulation lncRNAs
Supplementary Table S4. Analysis of SNP/INDEL
Supplementary Table S5. SNVs distribution of all samples
Supplementary Table S6. Statistics of AS events by JC
Supplementary Table S7. Statistics of AS events by JCEC
Supplementary Table S8. The list of DE lncRNAs
Supplementary Table S9. The list of DE mRNAs
Supplementary Table S10. Analysis of traits
Supplementary Figure S1. Statistics of Transcription factors (TFs). The abscissa is the TF family, the ordinate is the number of genes.
Supplementary Figure S2. Analysis of lncRNA type and the chromosome distribution of identified lncRNAs.
Supplementary Figure S3. Comparison of genomic characteristics between lncRNAs and mRNAs.
Supplementary Figure S4. The first 20 GO terms.
Supplementary Figure S5. Electropherogram analysis of Total RNA on Agilent 2100 Bioanalyzer.

REFERENCES

1. Shanshan W, Jianjun J, Zaiyan X, Bo Z. Functions and regul-

atory mechanisms of lncRNAs in skeletal myogenesis, muscle disease and meat production. Cells-Basel 2019;8:107. https://doi.org/10.3390/cells8091107

2. Li Y, Chen X, Sun H, Wang H. Long non-coding RNAs in the regulation of skeletal myogenesis and muscle diseases. Cancer Lett 2018;417:58-64. https://doi.org/10.1016/j.canlet.2017.12.015

3. Liu H, Xi Y, Liu G, Zhao Y, Li J, Lei M. Comparative transcriptomic analysis of skeletal muscle tissue during prenatal stages in Tongcheng and Yorkshire pig using RNA-seq. Funct Integr Genomics 2018;18:195-209. https://doi.org/10.1007/s10142-017-0584-6

4. Zhou Y, Liu S, Hu Y, et al. Comparative whole genome DNA methylation profiling across cattle tissues reveals global and tissue-specific methylation patterns. BMC Biol 2020;18:85. https://doi.org/10.1186/s12915-020-00793-5

5. Zillikens MC, Demissie S, Hsu YH, et al. Large meta-analysis of genome-wide association studies identifies five loci for lean body mass. Nat Commun 2017;8:80. https://doi.org/10.1038/s41467-017-00031-7

6. Gu H, Li J, Ying F, Zuo B, Xu Z. Analysis of differential gene expression of the transgenic pig with overexpression of PGC1alpha in muscle. Mol Biol Rep 2019;46:3427-35. https://doi.org/10.1007/s11033-019-04805-8

7. Zammit PS. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. Semin Cell Dev Biol 2017;72:19-32. https://doi.org/10.1016/j.semcdb.2017.11.011

8. Bunch H, Lawney BP, Burkholder A, et al. RNA polymerase II promoter-proximal pausing in mammalian long non-coding genes. Genomics 2016;108:64-77. https://doi.org/10.1016/j.ygeno.2016.07.003

9. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25:1915-27. https://doi.org/10.1101/gad.17446611

10. Luo W, Chen J, Li L, et al. C-Myc inhibits myoblast differentiation and promotes myoblast proliferation and muscle fibre hypertrophy by regulating the expression of its target genes, miRNAs and lincRNAs. Cell Death Differ 2019;26:426-42. https://doi.org/10.1038/s41418-018-0129-0

11. Li Y, Yuan J, Chen F, et al. Long noncoding RNA SAM promotes myoblast proliferation through stabilizing Sugt1 and facilitating kinetochore assembly. Nat Commun 2020;11:2725. https://doi.org/10.1038/s41467-020-16553-6

12. Cai B, Li Z, Ma M, et al. LncRNA-Six1 encodes a micropeptide to activate six1 in cis and is involved in cell proliferation and muscle growth. Front Physiol 2017;8:230. https://doi.org/10.3389/fphys.2017.00230

13. Wang L, Zhao Y, Bao X, et al. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Res 2015;25:335-50. https://doi.org/10.1038/cr.2015.21

14. Jin CF, Li Y, Ding XB, et al. Lnc133b, a novel, long non-coding RNA, regulates bovine skeletal muscle satellite cell proliferation and differentiation by mediating miR-133b. Gene 2017;630:35-43. https://doi.org/10.1016/j.gene.2017.07.066

15. Li D, Huang M, Zhuang Z, et al. Genomic analyses revealed the genetic difference and potential selection genes of growth traits in two duroc lines. Front Vet Sci 2021;8:725367. https://doi.org/10.3389/fvets.2021.725367

16. Yu J, Zhao P, Zheng X, Zhou L, Wang C, Liu JF. Genome-wide detection of selection signatures in duroc revealed candidate genes relating to growth and meat quality. G3 (Bethesda) 2020;10:3765-73. https://doi.org/10.1534/g3.120.401628

17. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-90. https://doi.org/10.1093/bioinformatics/bty560

18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9. https://doi.org/10.1038/nmeth.1923

19. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. Nat Methods 2015;12:357-60. https://doi.org/10.1038/nmeth.3317

20. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36:D480-4. https://doi.org/10.1093/nar/gkm882

21. Tafer H, Hofacker IL. RNAplex: A fast tool for RNA-RNA interaction search. Bioinformatics 2008;24:2657-63. https://doi.org/10.1093/bioinformatics/btn193

22. Lorenz R, Bernhart SH, Höner zu Siederdissen C, et al. Vienna RNA Package 2.0. Algorithms Mol Biol 2011;6:26. https://doi.org/10.1186/1748-7188-6-26

23. Wu Y, Wei B, Liu H, Li T, Rayner S. MiRPara: A SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. BMC Bioinformatics 2011;12:107. https://doi.org/10.1186/1471-2105-12-107

24. Der-Auwera GAV, Carneiro MO, Hartl C, Poplin R, Thibault J. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 2013;43:11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

25. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164. https://doi.org/10.1093/nar/gkq603

26. Ramaswami G, Zhang R, Piskol R, et al. Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 2013;10:128-32. https://doi.org/10.1038/nmeth.2330

27. Bahn JH, Lee JH, Li G, Greer C, Peng G, Xiao X. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res 2012;22:142-50. https://doi.org/10.1101/gr.124107.111

28. Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci 2014;39:170-82. https://doi.org/10.1016/j.tibs.2014.02.007

29. Ma M, Cai B, Jiang L, et al. LncRNA-Six1 is a target of miR-1611 that functions as a ceRNA to regulate six1 protein expression and fiber type switching in chicken myogenesis. Cells-Basel 2018;7:243. https://doi.org/10.3390/cells7120243

30. Chen LL. Linking long noncoding RNA localization and function. Trends Biochem Sci 2016;41:761-72. https://doi.org/10.1016/j.tibs.2016.07.003

31. Yotsukura S, Duverle D, Hancock T, Natsume-Kitatani Y, Mamitsuka H. Computational recognition for long non-coding RNA (lncRNA): Software and databases. Brief Bioinform 2017;18:9-27. https://doi.org/10.1093/bib/bbv114

32. Li R, Li B, Jiang A, et al. Exploring the lncRNAs related to skeletal muscle fiber types and meat quality traits in pigs. Genes (Basel) 2020;11:883. https://doi.org/10.3390/genes11080883

33. Jin W, Liu M, Peng J, Jiang S. Function analysis of Mef2c promoter in muscle differentiation. Biotechnol Appl Biochem 2017;64:647-56. https://doi.org/10.1002/bab.1524

34. Sun JY, Zhu ZR, Wang H, et al. Knockdown of UACA inhibitsproliferation and invasion and promotes senescence of hepatocellular carcinoma cells. Int J Clin Exp Pathol 2018; 11:4666-75.

35. Midwood KS, Chiquet M, Tucker RP, Orend G. Tenascin-C at a glance. J Cell Sci 2016;129:4321-7. https://doi.org/10.1242/jcs.190546

36. Yoshida T, Akatsuka T, Imanaka-Yoshida K. Tenascin-C and integrins in cancer. Cell Adh Migr 2015;9:96-104. https://doi.org/10.1080/19336918.2015.1008332

37. Zhou M, Li M, Liang X, et al. The significance of serum S100A9 and TNC levels as biomarkers in colorectal cancer. J Cancer 2019;10:5315-23. https://doi.org/10.7150/jca.31267

38. Simionescu-Bankston A, Kumar A. Noncoding RNAs in the regulation of skeletal muscle biology in health and disease. J Mol Med (Berl) 2016;94:853-66. https://doi.org/10.1007/s00109-016-1443-y

39. Li Y, Chen X, Sun H, Wang H. Long non-coding RNAs in the regulation of skeletal myogenesis and muscle diseases.

Cancer Lett 2018;417:58-64. https://doi.org/10.1016/j.canlet.2017.12.015

40. Tomida T, Adachi-Akahane S. Roles of p38 MAPK signaling in the skeletal muscle formation, regeneration, and pathology. Nihon Yakurigaku Zasshi 2020;155:241-7. https://doi.org/10.1254/fpj20030

41. Kumar A, Xie L, Ta CM, et al. SWELL1 regulates skeletal muscle cell size, intracellular signaling, adiposity and glucose metabolism. Elife 2020;9:e58941. https://doi.org/10.7554/eLife.58941

42. Aphasizhev R. RNA and DNA editing: methods and protocols. (Chapter 11):171-84. Humana Press; 2011.

43. Maas S, Kawahara Y, Tamburro KM, Nishikura K. A-to-I RNA editing and human disease. RNA Biol 2006;3:1-9.

44. Sarah, Djebali, Carrie, A., Davis, Angelika, et al. Landscape of transcription in human cells. Nature 2012;489:101-8.

45. Sebastian S, Faralli H, Yao Z, et al. Tissue-specific splicing of a ubiquitously expressed transcription factor is essential for muscle differentiation. Genes Dev 2013;27:1247-59. https://doi.org/10.1101/gad.215400.113

46. Shen S, Park JW, Lu Z, et al. RMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci USA 2014;111(51):E5593-601.

47. Hao Y, Feng Y, Yang P, et al. Transcriptome analysis reveals that constant heat stress modifies the metabolism and structure of the porcine longissimus dorsi skeletal muscle. Mol Genet Genomics 2016;291:2101-15. https://doi.org/10.1007/s00438-016-1242-8

48. Gennarelli M, Lucarelli M, Zelano G, Pizzuti A, Novelli G, Dallapiccola B. Different expression of the myotonin protein kinase gene in discrete areas of human brain. Biochem Biophys Res Commun 1995;216:489-94. https://doi.org/10.1006/bbrc.1995.2649

49. Zhang M, Zhu B, Davie J. Alternative splicing of MEF2C pre-mRNA controls its activity in normal myogenesis and promotes tumorigenicity in rhabdomyosarcoma cells. J Biol Chem 2015;290:310-24. https://doi.org/10.1074/jbc.M114.606277