Research Paper

Cis-Natural Antisense Transcripts Are Mainly Co-expressed with Their Sense Transcripts and Primarily Related to Energy Metabolic Pathways during Muscle Development

Yunxia Zhao1,2, Ye Hou1,2, Changzhi Zhao1,2, Fei Liu1,2, Yu Luan1,2, Lu Jing1,2, Xinyun Li1,2, Mengjin Zhu1,2,* Shuhong Zhao1,2,*

1. Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education and Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, 430070, PR China.
2. The Cooperative Innovation Center for Sustainable Pig Production, Wuhan, 430070, PR China

* Corresponding authors: Phone, 0086-2787281306; Fax, 086-02787280408; Email: zhumengjin@mail.hzau.edu.cn; Phone, 086-02787387480; Fax, 086-02787280408; Email: shzhao@mail.hzau.edu.cn. Address: College of Animal Sciences & Technology Huazhong Agricultural University, No.1, Shizishan Street Hongshan District, Wuhan, Hubei Province, P.R.China, 430070

© Ivyspring International Publisher. Reproduction is permitted for personal, noncommercial use, provided that the article is in whole, unmodified, and properly cited. See http://ivyspring.com/terms for terms and conditions.

Received: 2015.12.27; Accepted: 2016.04.30; Published: 2016.07.17

Abstract

Cis-natural antisense transcripts (cis-NATs) are a new class of RNAs identified in various species. However, the biological functions of cis-NATs are largely unknown. In this study, we investigated the transcriptional characteristics and functions of cis-NATs in the muscle tissue of lean Landrace and indigenous fatty Lantang pigs. In total, 3,306 cis-NATs of 2,469 annotated genes were identified in the muscle tissue of pigs. More than 1,300 cis-NATs correlated with their sense genes at the transcriptional level, and approximately 80% of them were co-expressed in the two breeds. Furthermore, over 1,200 differentially expressed cis-NATs were identified during muscle development. Function annotation showed that the cis-NATs participated in muscle development mainly by co-expressing with genes involved in energy metabolic pathways, including citrate cycle (TCA cycle), glycolysis or gluconeogenesis, mitochondrial activation and so on. Moreover, these cis-NATs and their sense genes abruptly increased at the transition from the late fetal stages to the early postnatal stages and then decreased along with muscle development. In conclusion, the cis-NATs in the muscle tissue of pigs were identified and determined to be mainly co-expressed with their sense genes. The co-expressed cis-NATs and their sense gene were primarily related to energy metabolic pathways during muscle development in pigs. Our results offered novel evidence on the roles of cis-NATs during the muscle development of pigs.

Key words: Cis-natural antisense transcript, muscle, energy metabolic pathway, pig

Introduction

Natural antisense transcripts (NATs) are endogenous RNAs that include cis- and trans-NATs. Cis-NATs are transcribed from the complementary DNA strand of annotated genes, whereas trans-NATs are transcribed from another region of the genome. Both NATs partially or totally overlap with their sense-annotated genes. In the past decade, cis-NATs have been identified in humans (approximately 30% of annotated genes) [1-3], mice (21.9%) [2, 3], pigs (approximately 30%) [4] and microorganisms [5, 6]. Approximately 7%~70% of the annotated genes of Arabidopsis have been associated with NATs [7-9]. These studies indicate that cis- and trans-NATs are widely expressed in eukaryotic species. Previous works suggested that cis-NATs were the by-products or noise of transcription [10, 11]; however, other
studies provided evidence on the roles of cis-NATs in gene expression regulation [12-15]. However, the biological functions of cis-NATs are still largely unknown because of the lack of systematic evaluation of their expression and sense genes.

Pioneer studies on eukaryotic identified four mainly antisense-mediated gene expression regulation mechanisms, including transcriptional interference, RNA masking, double-stranded RNA-dependent (dsRNA-dependent) mechanisms, and chromatin remodeling [16-18]. Transcriptional interference that could result in either transcription or transcription arrest in a single-direction and sense-antisense (SA) pair would be inversely expressed [14, 15, 19]. Under RNA masking, NATs could mask the splice sites on their sense pre-mRNA, which results in the generation of new isoforms of their sense genes [20-22]. Through RNA masking, NATs may be co-expressed with preferred splice variants and inversely expressed with repressed variants. In dsRNA-dependent mechanisms, including RNA editing and RNA interference, NATs are co-expressed with their sense genes [23, 24]. For chromatin remodeling, the transcription of NATs is involved in the expression of monoallelic genes or the silencing of nearby gene clusters [25-27]. The NATs and the genes in silenced clusters exhibit inverse expression under this mechanism. Determining the mechanism through which double-stranded RNA formed by the co-expressed NATs and their sense genes avoids degradation by endogenous RNAase may reveal a novel mechanism by which NATs regulate gene expression [28-30].

Skeletal muscle is a component of the motor system and also a major determinant of the resting energy expenditure of mammals [31]. Muscle development is a complex process that consists of four main stages: (i) differentiation of myogenic progenitors into myoblasts, (ii) fusion of myoblasts into myotubes, (iii) differentiation of myotubes into myofibers, and (iv) growth and maturation of myofibers [32-34]. The total number of fibers in pigs is largely determined by two major waves of muscle development: fetal stage 30-60-day period and 54-90-day period [33, 35, 36]. From birth to the first several weeks, myofibers differentiate into type I, IIA, IIX, and IIB muscle fibers [37, 38]. Type I muscle fibers are rich in mitochondria and enzymes that are involved in oxidative metabolism. Type IIB fibers exhibit the lowest levels of mitochondrial content and oxidative enzymes, whereas the oxidative functions of type IIA and IIX lie between types I and IIB [39, 40].

Previous studies revealed that cis-NATs are involved in the development of skeletal muscles. The cis-NATs of the MHC gene cluster (Myl1, Myh2, and Myh4) reportedly affect the expression patterns of different MHC genes in rats. The cis-NAT of Myh2 is significantly up-regulated in rat skeletal muscles after spinal cord isolation. Moreover, the expression of Myh2 cis-NAT negatively and positively correlates with the expression of Myh2 and Myh1, respectively, which could influence the composition of muscle fibers and meat quality [41]. Furthermore, the cis-NATs of PINK1 could target this protein coding gene to generate the PINK1 splice variant in human skeletal muscle tissue [20]. Recently, whole genome studies have revealed that cis-NATs widely exist in the muscle tissue of human [1] and pig [4]. However, the expression characteristics, biological functions, and regulatory roles of cis-NATs in skeletal muscle development remain unclear. In the present study, we conducted a genome-wide identification of cis-NATs in muscle tissue by using the digital gene expression (DGE) profiling libraries of Lantang (LT) and Landrace (LR) pigs and the dUTP RNA-seq data of Yorkshire (YS) pigs. We also systematically analyzed the expression characteristics and functions of cis-NATs during the muscle development of pigs.

Materials and Methods

DGE data download and dUTP sequencing

The clean data of 20 DGE libraries with series ID GSE25406 were downloaded from Gene Expression Omnibus. This DGE data were generated by sequencing the longissimus skeletal muscle samples of LT and LR at fetal days 35, 49, 63, 77, and 91 and postnatal days 2, 28, 90, 120, and 180 [42]. Moreover, the total RNA of the longissimus muscle tissues of six 180-day old YS pigs was isolated using the TRIzol reagent (Invitrogen). Equal RNA of each sample was used for dUTP strand-specific sequencing library construction. The second-strand marking mix was employed to replace dTTP with dUTP during second-strand cDNA synthesis. RNA sequencing was performed by a commercial service in Genery Biotechnology Company (Shanghai, China). All animal experimental protocols were approved by the Ethics Committee of Huazhong Agricultural University (HZAUMU2013-0005).

Sequence data alignment and annotation

All sequencing data were mapped to the pig genome (Sscrofa10.2) by utilizing bowtie 0.12.7 [43]. Up to one and two mismatches were allowed for the unique mapping of the DGE and dUTP data, respectively. The unmapped tags of the DGE data were also mapped to the Sscrofa RNA reference, and one mismatch was allowed. The genome-mapped data were annotated using the gff3 file of the Sscrofa10.2 genome (ftp://ftp.ncbi.nlm.nih.gov/
Genome-wide identification of SA genes

The copy number of raw clean tags of DGE data was normalized to tags per million (TPM) [42]. The TPM of tags that belonged to up two transcripts were equally divided. The expression level of the sense and antisense strands of annotated genes was summed by the TPM of sense exon-mapped tags and all antisense tags, respectively. The SA genes that were identified with the TPM of the sense and antisense strands were $\geq 5$ and $\geq 2.5$, respectively.

The six dUTP libraries were pooled for further analysis. The expression level of both strands of annotated genes was calculated using the same method as that of the DGE data. The SA genes of the YS pigs identified on the basis of the read number of the sense and antisense strands must be $\geq 30$ and 5, respectively.

Genome-wide identification of cis-NAT regions

For DGE library construction, NlaIII (5′-CATG-3′) was used to digest the cDNA. Then, cDNA fragments with a Poly A tail were isolated using Poly T-coated beads. The adapter with the Mmel enzyme site was ligated with the isolated cDNA fragments. Finally, 21 bp tags were generated for sequencing through digestion with the Mmel enzyme. The sense and antisense tags could be distinguished on the basis of the cleaving site sequence of NlaIII (Supplementary Figure S1). To avoid a false positive, we further selected the antisense tags for cis-NAT identification on the basis of the following criteria: (i) the tags should be mapped on the antisense strands of the annotated genes. (ii) The TPM of tags should be at least two. (iii) The TPM of tags should be higher than their corresponding sense tags or 5′ upstream tags on the same strands. The antisense tags in each DGE library were identified independently.

The cis-NAT region was finally identified through the following steps: (i) the neighboring antisense tags of the DGE data were clustered as one cis-NAT region when the TPM of the 3′ terminal tag was higher than that of the 5′ terminal tag. (ii) The overlapped cis-NAT regions that were identified from different DGE libraries were merged into one large cis-NAT region. (iii) For the dUTP data, the cis-NAT regions were identified directly by assembling the antisense sequence fragments. (iv) The overlapped cis-NAT regions from the DGE and dUTP data were merged again to determine the final cis-NAT regions.

Correlation analysis of cis-NATs and their sense genes

The expression profiling of cis-NATs was calculated first. The antisense tags that significantly positively ($P < 0.1$) correlated with their sense tags were removed to avoid the interference of sense tags. The expression profiling of cis-NATs was summed by the remaining antisense tags within the cis-NAT regions.

The Pearson correlation between the cis-NAT and sense gene pairs (total TPM of one of the pair $\geq 5$ and detected stages $\geq 3$) was analyzed. The P value was adjusted by the BH method, and the significance level was FDR < 0.05. Range analysis was also performed to investigate the expression pattern between the cis-NAT and sense gene pairs. cis-NAT and sense gene pairs were selected for the range analysis on the basis of the following criteria: For both cis-NAT and its sense genes, TPM > 2 stages must be at least four, and the total TPM should be at least 10. The fold change (FC) between the sense gene and its cis-NAT was calculated at each stage, and the maximum and minimum FC stages were identified. The correlated cis-NAT and sense gene pairs were determined when their FC between these two stages was higher than 1.5. The co-expressed pairs were determined when their change direction of FC was the same between these two stages. Otherwise, they were identified as inversely expressed pairs.

Correlation analysis of cis-NATs and the isoforms of sense genes

The isoforms of the SA genes were identified by utilizing Tophat 2.0.9 [44] and Cufflinks [45]. The expression level of each isoform was defined using the fragments per kilo base of exon per million fragments mapped (FPKM) in accordance with the following formula [46]:

$$ FPKM = \frac{\text{total reads mapped to genome } \times \text{gene length} \times 10^6}{\text{total reads}} $$

The correlation between the cis-NATs and their sense gene isoforms was detected through Pearson correlation analysis. The significant threshold was set at $|R| > 0.75$.

Identification of differentially expressed transcripts

The differential expression of the annotated genes and cis-NATs among different stages was identified by employing edgeR [47]. The read counts of the annotated genes and cis-NATs were normalized by edgeR. The criterion of significantly different expression was $|\log_{2}FC| \geq 1$ and FDR < 0.01.

qPCR was performed to validate the
differentially expressed (DE) sense genes and their cis-NATs. The primers are listed in Supplementary Table S2. qPCR was carried out on a BIO-RAD CFX384 Real-Time System using SYBR Green PCR Master Mix Kit (BIO-RAD, America). The expression levels of the genes were analyzed by adopting the 2-ΔΔCt method, and the 18S gene was employed as the internal control. Student's t-test was used to analyze the expression difference between postnatal day 2 and day 180.

Functional annotation

The significant biological process GO terms (P < 0.01) and pathways (P < 0.05) of the DE annotated genes and cis-NATs were enriched via DAVID Bioinformatics Resources 6.7 [48]. Significant pathways containing less than five correlated cis-NAT and sense gene pairs were discarded. The distribution of the correlated cis-NAT and sense gene pairs in these significant pathways during muscle development was investigated. Genes involved in energy metabolic pathways and their correlated cis-NATs were visualized using Cystoscape [49].

Results

A total of 3,561 SA genes have been identified based on the data of 20 DGE libraries

To identify the cis-NATs, we identified the genes with antisense transcripts first, which were defined as SA genes. A total of 65.06 million unique mapping tags were identified from 20 DGE libraries. Among these tags, 47.96 million (73.72%) were mapped onto annotated genes and 6.18 million (9.49%) were mapped onto their antisense strands. These antisense strand-mapped tags were further used to detect the SA genes and cis-NATs. Moreover, the dUTP sequencing data of six YS pigs were used to confirm the results of the SA genes and cis-NATs identified from the DGE data. A total of 33.60 million uniquely mapped reads were found in the dUTP data. Among these mapped reads, 22.02 million (65.52%) and 599,047 (1.78%) were mapped onto the sense and antisense strands of the annotated genes, respectively.

A total of 3,561 and 2,862 SA genes were identified using the DGE and dUTP data, respectively (Supplementary Table S3). Moreover, 1,182 SA genes were identified in both DGE and dUTP data. Fisher's exact test showed that the overlap between these two types of data was significant (P < 0.01) (Supplementary Figure S2 and Supplementary Table S4).

In total, 3,306 cis-NAT regions have been identified

The cis-NATs were further identified on the basis of the results of the SA genes. First, the characteristics of the mapped tags of the annotated genes were analyzed. A total of 54.6% of the annotated genes contained only one tag, and the others had more than one tag (TPM ≥ 2) (Supplementary Figure S3). For the later genes, the TPM of 83.68% of the 5′ terminal tags was lower than that of the 3′ terminal tags (Supplementary Figure S3). Therefore, the neighboring tags could be clustered when the TPM of the 5′ terminal tags was lower than that of the 3′ terminal tags. Statistical results demonstrated that approximately 80% of the tags were clustered (Supplementary Figure S3). Among these tags, 37.6% and 40.4% of the sense genes contained one cluster and two clusters, respectively; the other sense genes contained at least three clusters (Supplementary Figure S3). The results of expression correlation analysis indicated that the clusters and full-length annotated genes were significantly positively correlated (P < 0.01, R > 0.5) (Supplementary Table S5).

To identify the cis-NATs, the antisense tags of the annotated genes were also clustered. A total of 3,523 cis-NAT regions of the 2,469 SA genes were identified. Furthermore, 892 SA genes did not detect any cis-NAT region because of a single tag or the TPM of the 5′ terminal tags was higher than that of the 3′ terminal tags. The cis-NAT regions were also confirmed by using dUTP data. Based on that, 1,081 of the 3,523 identified cis-NATs were merged into 864 cis-NATs. Moreover, statistical results showed that 56.94% (492) of these merged cis-NATs were longer than 1 kb (Supplementary Figure S3). Finally, 3,306 cis-NAT regions of 2,469 SA genes were identified (Supplementary Data 1: gff file of cis-NATs).

Validation of Cis-NATs

RT-PCR was performed to further confirm the cis-NATs. First, the DNA and RNA samples were isolated from the longissimus muscle tissues of 2-, 40-, and 180-day old YS pigs. Six cis-NATs and their sense genes were randomly selected for validation. Two pairs of primers were specifically designed for each of the SA genes. One primer pair was designed for sense transcription amplification, which was bound to different exons and could be used to amplify both cDNA and DNA templates. The size of the PCR products of the cDNA template was less than that of the DNA template. Thus, only one band was amplified for each cDNA sample, and none of the cDNA samples was contaminated with DNA. The other primer pair was designed for cis-NAT amplification. For the primers of each cis-NAT, at least one primer was designed to bind to the intron regions of its sense gene. Therefore, the sense gene could not
be amplified by the primers of cis-NATs. PCR results showed that all the six cis-NATs could be amplified, and the cis-NAT band was similar in size to the DNA template band (Supplementary Figure S4). In addition, 100 sense genes related to muscle growth were randomly selected in accordance with the GO terms (Supplementary Data 2), and 45 sense genes were confirmed to have antisense transcripts on the basis of the latest annotation of the human, mouse, or pig genome.

**Cis-NATs were primarily expressed at fetal stages and sharply decreased after birth**

The expression patterns of the SA genes and cis-NATs were further investigated. The number of annotated genes and SA genes at the fetal stages was obviously higher than that at the postnatal stages (Figure 1A). A sharp decrease in these expression patterns was observed during the transition of fetal day 91 to postnatal day 28 (Figure 1A). The distributions of the highest and lowest expressed cis-NATs during the 10 stages were statistically analyzed. The highest expressed cis-NATs were mainly distributed at the fetal stages, and the lowest expressed cis-NATs were mainly distributed at the postnatal stages. Extreme changes occurred during the transition of fetal day 91 to postnatal day 28 (Figure 1B). To compare and analyze further the expression of cis-NATs between the fetal and postnatal stages, the cis-NATs were classified into four categories: (i) The highest and lowest expression levels appeared at the fetal stages (fetal–fetal); (ii) The highest and lowest expression levels appeared at the postnatal stages (postnatal–postnatal); (iii) The expression level was the highest at the fetal stages and the lowest at the postnatal stages (fetal–postnatal); (iv) The expression level was the highest at the postnatal stages and the lowest at the fetal stages (postnatal–fetal). Statistical results showed that more than 50% of the cis-NATs belonged to the fetal–postnatal category (Figure 1C). These results indicate that the cis-NATs were highly expressed at the fetal stages and sharply decreased after birth.

**Figure 1.** Expression pattern of cis-NATs during muscle development. Letters F and P on the x-axis indicate fetal and postnatal, respectively. LR and LT represent the Landrace and Lantang pig breeds, respectively. (A) Number of expressed annotated genes and SA genes at each stage. Blue triangle (Δ), red dot (○), and grey diamond (◊) indicate LR pigs, LT pigs, and mean value of LR and LT pigs, respectively. (B) Distribution of the highest and lowest expressed cis-NATs. The number of the highest and lowest expressed cis-NATs at each stage was represented by a green triangle (Δ) and orange dot (○), respectively. (C) Statistical results of the four types of cis-NATs. Fetal–fetal type means that both the highest and lowest expression levels of cis-NATs occurred at the fetal stages. Postnatal–postnatal type denotes both conditions occurred at the postnatal stages. Fetal–postnatal type means that the highest expression occurred at the fetal stages, whereas the lowest expression occurred at the postnatal stages. Postnatal–postnatal type represents the opposite of the fetal–postnatal type.
Cis-NATs and their sense genes were co-expressed

In total, 1,039 cis-NAT and sense gene pairs in LR and 1,144 in LT significantly correlated accordance with Pearson correlation analysis (|R| > 0.70 and FDR < 0.05) (Table 1). Moreover, 656 significantly correlated cis-NAT and sense gene pairs were detected in both LR and LT pigs. Fisher’s exact test showed that the overlap was significant (P < 0.01) (Supplementary Table S6). The correlation between the expression levels of cis-NATs and their sense genes was also analyzed using the range analysis method. A total of 672 (LR) and 682 (LT) correlated cis-NAT and sense gene pairs were identified (Table 1). Moreover, 230 correlated pairs were determined in the two breeds, and the overlap was significant (P < 0.01) (Supplementary Table S7). The correlated pairs identified by these two methods was 0.93 (Figure 3E).

### Table 1. Statistical analysis of expression correlation between sense genes and their cis-NATs

| Analysis        | Breed | Correlation | Total | Overlap |
|-----------------|-------|-------------|-------|---------|
|                 |       | Pearson     |       |         |
|                 |       | (FDR < 0.05) | Positive/Co-expression (%): 1,019 (98.08%) | 20 (1.92%) | 1,039 | 656** | 63.14% |
|                 |       | Range analysis | (log2FC ≥ 1) | 1,138 (99.48%) | 6 (0.52%) | 1,144 | 57.34% |

The correlation between cis-NATs and the isoforms of their sense genes was further analyzed. A total of 384 SA pairs of 355 SA genes exhibited significant (R > 0.75 or R < −0.75) correlations (Supplementary Data 4), and 55.99% positively correlated (Figure 2C left bar plot). Furthermore, 200 SA genes with at least two expressed isoforms were selected to detect the correlation between the cis-NATs and specific isoforms (Figure 2C right bar plot). The correlation between the cis-NATs and specific isoforms of the SA genes was observed. The gene structure and specific correlations in the HEB2 and COPB2 SA genes, as two typical illustrations, are shown (Figure 2D and Supplementary Figure S2). For these two genes, the partial alternative splicing sites were masked by their cis-NATs. A strong positive correlation (R=0.9) existed between cis-NAT and the dominant expressed isoform NM_213921.1 / TCONS_00017836 of the HEB2 SA gene, and a similar positive correlation (R=0.85) was observed in the COPB2 SA gene (Figure 2D and Supplementary Figure S2). Moreover, the cis-NAT was moderately and negatively correlated with another isoform TCONS_00017835 of the HEB2 SA gene (R = −0.74).

Weak positive and negative correlations were also observed between the cis-NAT and the other isoforms of the COPB2 SA gene (Supplementary Figure S2). These results support the hypothesis that cis-NATs could affect the alternative splicing and co-expressed with their preferred isoforms of the sense genes.

Over 1,200 DE cis-NATs were identified during muscle development

DE cis-NATs and annotated genes were also analyzed. During the 10 stages of muscle development, 9,209 (LR) and 8,860 (LT) DE annotated genes were identified (log2FC ≥ 1 and FDR < 0.01) (Supplementary Table S8 and Supplementary Data 5). Moreover, 1,281 (LR) and 1,211 (LT) DE cis-NATs were detected (Supplementary Table S8 and Supplementary Data 5). The divergence of the DE cis-NATs and annotated genes was calculated through systematic cluster analysis (Figure 3). The divergence of the annotated genes and cis-NATs was similar in the two breeds. All of the fetal stages were clustered together, as well as the postnatal stages. These results indicate that the expression patterns of annotated genes and cis-NATs at the fetal stages significantly differed at the postnatal stages. In addition, three important transitions, namely, F35 to F49, F91 to P28 d., and P90 to P120 d, were determined on the basis of the number of DE cis-NATs (Supplementary Table S8).

Furthermore, the DE cis-NATs and annotated genes were validated by the qPCR method. Five transcriptions, namely, two DE cis-NATs (MYBPC1-AS1 and ENO3-AS1, FDR<0.05), one un-DE cis-NAT (COX6A-AS1, FDR=0.08), and two DE annotated genes (COX6A and ENO3, FDR<0.05), between postnatal day 2 and day 180 were selected for qPCR validation (Supplementary Data 5). The qPCR results were similar to the RNA-seq results, and the correlation coefficient of the FC values between the two methods was 0.93 (Figure 3E).
Figure 2. Identification of correlated cis-NAT and sense gene pairs. SA, LR, and LT indicate sense–antisense, Landrace, and Lantang pigs, respectively. (A) Number of correlated cis-NAT and sense gene pairs. (B) Venn plot of the correlated cis-NAT and sense gene pairs. The correlated cis-NAT and sense gene pairs were analyzed using both Pearson correlation and range methods. (C) Statistical analysis of the positively and negatively correlated SA pairs (left bar plot) and the three types of SA genes (without any isoforms, with unexpressed isoforms, and with more isoforms) (right bar plot) in the muscle tissue of adult YS pigs. The SA pairs under the left bar indicate the isoforms of the SA gene and its cis-NAT pairs. (D) Illustration of cis-NATs related to the alternative splicing of their sense genes. FPKM and TTS denote fragments per kilobase of exon per million fragments mapped and transcription termination site, respectively. (E) Four co-expressed cis-NAT and sense genes were detected using qPCR.
Cis-NATs participated in muscle development by co-expressing with the annotated genes of energy metabolic pathways in pigs

The functional annotation of the cis-NATs in muscle development was based on their correlated annotated genes. The enriched pathways of the DE annotated genes or sense genes of DE cis-NATs were analyzed using DAVID. The top 16 significant KEGG pathways were identified on the basis of the number of correlated cis-NAT and sense gene pairs. Eight of the 16 pathways were involved in energy metabolism, including TCA cycle, fatty acid metabolism, propanoate metabolism, pyruvate metabolism, glycolysis or gluconeogenesis, oxidative phosphorylation, Alzheimer’s disease, Huntington’s disease, and Parkinson’s disease (Figure 4). The eight other pathways were involved in myoblast differentiation and muscle growth, which included arginine and proline metabolism, hypertrophic cardiomyopathy, cardiac muscle contraction and so on (Figure 4). Moreover, the results of GO term enrichment demonstrated that the bioprocesses of energy metabolism and muscle development could be
regulated by the correlated cis-NAT and sense gene pairs (Supplementary Data 6).

The expression details of the genes and their correlated cis-NATs were further studied. Most of the genes were involved in energy metabolic pathways, and their cis-NATs were gradually up-regulated from fetal day 49 to postnatal day 2 in the two breeds (Figure 5A and Supplementary Data 7). Furthermore, these genes and their correlated cis-NATs were abruptly down-regulated during the transition from postnatal day 2 to day 28 in the LR pigs (Figure 5C and Supplementary Data 7) and from postnatal day 28 to day 90 in the LT pigs (Supplementary Figure S3 and Supplementary Data 7). These results indicate that cis-NATs participated in muscle development by co-expressing with the genes involved in energy metabolic pathways.

Discussion

In this study, cis-NATs were first identified during muscle development. The expression patterns and functions of these cis-NATs were thoroughly analyzed. Results showed that the cis-NATs were mainly co-expressed with their sense genes and were mainly related to energy metabolic pathways during muscle development. In the past decade, studies on cis-NATs mainly focused on identification [3, 7, 8, 50], conservation analysis [1, 3, 50, 51], or tissue-specific expression detection [7, 50]. Previous studies suggested that cis-NATs were transcriptional noise [10, 11], whereas other studies have confirmed that cis-NATs are pervasive in various species [3, 7, 8, 50] and regulate gene expression [12-15]. A previous work reported that 33.7% annotated genes have cis-NATs in the muscle of adult pigs and considered 226 cis-NATs to be cis-antisense eQTLs (expression quantitative trait loci) [4]. However, the expression characteristics and roles of cis-NATs during muscle development remain to be studied comprehensively.

Therefore, this study offered new evidence on the roles of cis-NATs during muscle development in pigs. We systematically analyzed the expression patterns of cis-NATs during the muscle development of pigs. In general, the number of cis-NATs expressed at the fetal stages was higher than that at the postnatal stages. These results indicate that cis-NATs are mainly expressed at the fetal stages and play important roles during muscle development. Moreover, approximately 40% of the cis-NATs correlated with their sense genes at the transcriptional level, and about 80% of them were co-expressed. Previous studies reported that NATs and their sense genes are co-expressed in human HeLa cells [28] and D. melanogaster S2 cells [31]. In humans, 672 of the 1,498 cis-NATs and their sense genes were identified to be co-expressed on the basis of the expression profile analysis in 49 tissues [1]. The IGF2 cis-NAT IGF2-AS1 is reportedly lost in imprinting and co-expressed with IGF2 in human Wilms’ Tumor [30, 52].

![Figure 4](http://www.ijbs.com)

**Figure 4.** Heat map of the top 16 enriched pathways. LR, Landrace; LT, Lantang. (A) Heat map of the top 16 enriched pathways in LR pigs. (B) Heat map of the top 16 enrichment pathways in LT pigs. The significant pathways were enriched using DAVID tool and ranked on the basis of the number of correlated sense gene and cis-NAT pairs they have.
Figure 5. Expression patterns in energy metabolic pathways. The expression details of genes involved in energy metabolic pathways and their correlated cis-NATs are illustrated. (A–B) Expression patterns of transition fetal day 91 to postnatal day 2 and postnatal day 2 to postnatal day 28. Green, blue, pink, and red fill colors represent log₂ FC ≤ -0.5, -0.5 < log₂ FC ≤ 0, 0 < log₂ FC < 0.5, and log₂ FC ≥ 0.5, respectively. LR, Landrace.
These results indicate that the expression of cis-NATs positively correlated with the expression of their sense genes. This finding exceeded our expectation because we assumed that the cis-NATs would negatively correlate with the expression of their sense genes on the basis of the mechanism of RNAi. However, the mechanisms underlying the co-expression between the cis-NATs and their sense genes remain largely unknown. One of the possible explanations is that the cis-NATs and their sense gene pairs could escape the dicer-associated RNAi pathway [28]. Another possibility is that the cis-NATs could mask the specific sites of the RNAs of the sense genes and maintain their stability [22, 53-55]. In the present study, the cis-NATs could positively correlate with the specific isoforms of their sense genes. This finding supports that cis-NATs could mask the alternative splicing sites and promote the expression of the specific isoforms of their sense genes.

The SA genes that were correlated with the cis-NATs were mainly involved in energy metabolic pathways, including TCA cycle, glycolysis or gluconeogenesis, fatty acid metabolism, pyruvate metabolism, propanoate metabolism, and mitochondrial activation pathways. In these pathways, the respiration chain complex subunits encoding genes NDUFA4, NDUFB3, UQRC1, ATP5J, and COX7A2L and other important energy metabolic genes IDH2 [56], DLD [57], PINK1 [20] and, SUCLG2 [58, 59] were determined to have co-expressed cis-NATs. Therefore, we deduced that cis-NATs play important roles in muscle development by co-expressing with genes involved in energy metabolic pathways. Moreover, all the co-expressed energy metabolic-related genes and their cis-NATs sharply increased from day 6 to postnatal day 2 in the LR and LT pigs. After birth, the cis-NATs were sharply down-regulated from postnatal day 2 to day 28 in the LR pigs. Meanwhile, the expression profile of the cis-NATs was almost unchanged from day 2 to day 28 but sharply down-regulated from postnatal day 28 to day 90 in the LT pigs. Previous studies indicated that secondary muscle fibers begin to form around fetal day 60 and last until the end of the fetal stages; meanwhile, type II muscle fibers begin to form at the early postnatal stages [38, 43]. Thus, we deduced that the cis-NATs could be involved in the development of muscle fibers in pigs. Furthermore, previous studies determined that the proportion of type II muscle fibers is higher in LR than that in LT pigs at the early postnatal stages [60, 61]. This occurrence was coincidental with the earlier down-regulation of energy metabolic pathways involving genes and their correlated cis-NATs in the LR pigs at the early postnatal stages. Thus, cis-NATs may also relate to muscle fiber proportion at the early postnatal stages in pigs.

Supplementary Material

**Additional File 1:**
Supplementary figures and tables.
http://www.ijbs.com/v12p1010s1.pdf

**Additional File 2:**
Supplementary Data 1.
http://www.ijbs.com/v12p1010s2.xls

**Additional File 3:**
Supplementary Data 2.
http://www.ijbs.com/v12p1010s3.xls

**Additional File 4:**
Supplementary Data 3.
http://www.ijbs.com/v12p1010s4.xls

**Additional File 5:**
Supplementary Data 4.
http://www.ijbs.com/v12p1010s5.xls

**Additional File 6:**
Supplementary Data 5.
http://www.ijbs.com/v12p1010s6.xls

**Additional File 7:**
Supplementary Data 6.
http://www.ijbs.com/v12p1010s7.xls

**Additional File 8:**
Supplementary Data 7: video A.
http://www.ijbs.com/v12p1010s8.wmv

**Additional File 9:**
Supplementary Data 7: video B.
http://www.ijbs.com/v12p1010s9.wmv

Acknowledgment

This research was supported by an NSFC Outstanding Youth grant (31025026), and grants from the National Key Basic Research Program of China (2012CB124702) and National Natural Science Foundation of China (31372291).

Authors’ contributions

XL, MZ and SZ conceived and supervised the project. YZ, FL, YL and LJ undertook the analysis. YZ wrote the paper. CZ and YH performed the PCR experiments. All authors read and approved the final manuscript.

Competing Interests

The authors have declared that no competing interest exists.

References

1. Chen J, Sun M, Hurst LD, et al. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet. 2005; 21: 326-9.
2. Katayama S, Tomaru Y, Kasukawa T, et al. Antisense transcription in the mammalian transcriptome. Science. 2005; 309: 1564-6.
