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

Authors

Yunxia Zhao[1,2], Ye Hou[1,2], Changzhi Zhao[1,2], Fei Liu[1,2], Yu Luan[1,2], Lu Jing[1,2], Xinyun Li[1,2] , Mengjin Zhu[1,2,*], Shuhong Zhao[1,2,*]

* Corresponding author

Phone number, 0086-2787281306; Fax number, 086-027-87280408; Email addresses: zhumengjin@mail.hzau.edu.cn

Phone number, 086-027-87387480; Fax number, 086-027-87280408; Email addresses: 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

Affiliations

[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

Supplementary Tables

Table S1. RT-PCR primers of six selected *cis*-NATs and their sense genes

Number	Symbol	Primer sequence 5′–3′	Product length (bp)		Tm (℃)
			cDNA	DNA	
1	*COL12A1*	F: CCGGGCGAACTTTAGAACTG R: GCTCCACGCCTTCATCCTTC	354	537	60
	COL12A1-AS1	F: TCTGCCCCCAAAGAACCAAA R:GGAAAGGTCAGGCTGGAGAG	561	561	60
2	*PRKG1*	F: CGTGGATACAAGACAGCAGGA R:GATCCCTGAGAATGGTCCAGA	166	1001	60
	PRKG1-AS1	F: AAGCCCCACCTTCCTACACA R:TGCTGGGACAGTTTTAGGGTC	519	519	58
3	*MYL1*	F: TTGTCAAGCACATCATGTCTATCTA R:GAGCAGCAGACACTTGGTTT	234	806	60
	MYL1-AS1	F: TGTCAAAGGGGGTGGTCAAT R:CCAGGAGCACACTTACCCTC	455	455	60
4	*MYOZ1*	F: ATTCTCCTACAGCAAGGGCAG R:CAGCAGATCAATGCCAAGCTC	306	667	60
	MYOZ1-AS1	F: CCCATTGCTACCTTCCCCTC R:GGGAAGTTGCTAGGTCCAGG	564	564	60
5	*RYR1*	F: AGCTGAACGAGTACAATGCCT R:TCGTCGATGCCCAAGTTGTT	366	616	60
	RYR1-AS3	F: TCTGCTGGTTAGGGAAGGGG R:TCTCCACTGCATGGTCACCT	525	525	60
6	*MYBPC1*	F: GTGAAGCAGCAGGAGGAAGA R:TGCCAGTTCCACAGCAAATCT	236	644	60
	MYBPC1-AS1	F: CCTCTTCTCTGTGGCAGGTG R:CGGCAAAGAGGAAACACAGC	663	663	60

F, Forward; R, Reverse

Table S2. qPCR primers of four co-expressed *cis*-NAT and sense gene pairs

Number	Symbol	Primer sequence 5′–3′	Product length (bp)	Tm (℃)
1	*MYBPC1*	F: AAGAATGCGAACCCCAACGAG R: ACCCTGCCTCCTTTATCAACCTG	167	60

	MYBPC1-AS1	F: TTGTTTGGTCTCTTCCGTCA R: TAGGTCATATGTCGCCTCC	199	60
2	*COX6A1*	F: ATCTTCGCATCAGGTCCAAGC R: TTCATCTTCATAGCCAGTTGGAA	101	60
	COX6A1-AS1	F: CTGTTCCTCTCCTTTACGCTA R: AGAAACCTTGCTGTATGGCTT	178	60
	ENO3	F: GCCCAGCGAAGACATCCCA R: CAGCTCGGAATCGGCCCTT	122	60
3	*ENO3-AS1*	F: CCCTTCTCAGGCCTTCAATGT R: GTCCTTCCCGTATTTGCCCTT	175	60
	ACVR2A	F: CTGCCATATCTCACAGGGACA R: CAACCTGCCCATGGGTATCA	135	60
6	*ACVR2A-AS1*	F: GCCAGGTAGAGAATCCCTGA R: AGGCTTTAAGAGGTCTAGCCA	116	60

Table S3. Genome-wide identification of SA genes in porcine skeletal muscle

	Libraries	SA genes [b]	Overlap	
			Count	Percentage
DGE libraries [a]	20	3,561	1,182	33.19%
dUTP libraries	6	2,862		34.86%

[a] DGE, Illumina's Digital Gene Expression Profiling;

[b] SA, sense-antisense.

The SA genes at the third column were identified by using DGE or dUTP RNA-seq data. The count of overlapped SA genes (the fourth column) was the same as that of SA genes that were identified in both the DGE and dUTP libraries. The percentage of the overlapping in each type of library (the fifth column) was calculated by dividing the number of overlapped SA genes by the number of identified SA genes in each type of library.

Table S4. 2×2 contingency table of two types of libraries

	dUTP libraries	DGE libraries [c]	P value
Non-SA genes [a]	9,753	2,209	< 0.01
SA genes [b]	2,862	1,182	

[a] Non-SA genes (the 1st row), genes detected by RNA-seq but not detected to have *cis*-NATs.

[b] SA, sense–antisense.

[c] DGE, Illumina's Digital Gene Expression Profiling;

In dUTP libraries, 12,615 genes were detected to be expressed, of which 2,862 were identified as SA genes. In DGE libraries, 3,391 (the sum of the third column) of 3,561 SA genes were detected to be expressed in dUTP libraries and, among them, 1182 were also identified as SA genes in dUTP libraries. The P value was calculated by Fisher's exact test, which was performed to analyze the overlap of SA

genes between DGE and dUTP libraries.

Table S5. Correlation of expression profiling between fragments and the full length gene

Breed	R	P value
LT[a]	0.59	< 0.01
LR[b]	0.53	< 0.01

[a] LR, Landrace;

[b] LT, Lantang.

Table S6. 2×2 contingency table of significantly correlated *cis*-NAT and sense genes pairs from LR and LT pigs (Pearson correlation analysis)

	LR [a]	LT [b]	P value
Not significantly correlated	1,656	401	< 0.01
Significantly correlated	1,039	656	

[a] LR, Landrace;

[b] LT, Lantang.

For the LR pigs, 2,695 (sum of the second column) *cis*-NAT and sense gene pairs were used to carry out correlation analysis by expression profiling filtering, and 1,039 of the pairs were significantly (FDR < 0.05) correlated. For the LT pigs, 1144 significantly (FDR < 0.05) correlated pairs were identified, among them 656 pairs were also determined as significantly correlated pairs in the LR pigs, and the 87 remaining significantly correlated pairs of the LT pigs were not included in the LR pigs. Fisher's exact test was carried out to calculate the P value of the overlap of significantly correlated pairs between the DGE and dUTP libraries.

Table S7. 2×2 contingency table of correlated *cis*-NAT and sense gene pairs from LR and LT pigs (Range analysis)

	LR [a]	LT [b]	P value
Not correlated	933	192	P < 0.01
Correlated	674	230	

[a] LR, Landrace;

[b] LT, Lantang.

We first discarded the low-expression profiling *cis*-NAT pairs that were described in the materials and methods part. For the LR pigs, 1,607 sense and *cis*-NAT pairs were used to perform range analysis, and 674 of the *cis*-NATs pairs were identified as correlated pairs. For the LT pigs, 586 correlated pairs were identified, and 230 of the correlated pairs were also determined in LR pigs. The 68 remaining correlated pairs of the LT pigs were not included in the 1,607 *cis*-NAT and sense gene pairs that were

used for the Range analysis in the LR pigs. The P value was evaluated by Fisher's exact test on the overlap of correlated *cis*-NAT and sense gene pairs.

Table S8. Summary of differentially expressed transcripts during the muscle development of LR and LT pigs

Comparison	Numbers of differentially expressed transcripts			
	Sense	**Sum**	***Cis*-NATs**	**Sum**
LR-F49d / F35d	2,259		274	
LR-F63d / F49d	376		45	
LR-F77d / F63d	146		25	
LR-F91d / F77d	373		60	
LR-P2d / F91d	2,083	9,209	295	1,281
LR-P28d / P2d	1,124		137	
LR-P90d / P28d	661		111	
LR-P120d / P90d	1,433		221	
LR-P180d / P120d	754		113	
LT-F49d / F35d	1,342		169	
LT-F63d / F49d	701		96	
LT-F77d / F63d	268		34	
LT-F91d / F77d	252		51	
LT-P2d / F91d	2,587	8,860	314	1,211
LT-P28d / P2d	1,527		155	
LT-P90d / P28d	1,180		220	
LT-P120d / P90d	681		114	
LT-P180d / P120d	322		58	

LR, Landrace;

LR-F or LR-P, Landrace-fetal or Landrace-postnatal;

LT, Lantang;

LT-F or LT-P, Lantang-fetal or Lantang-postnatal;

Sense, differentially expressed transcripts of annotated genes;

Cis-NATs, differentially expressed *cis*-nature antisense transcripts.

Supplementary Figures

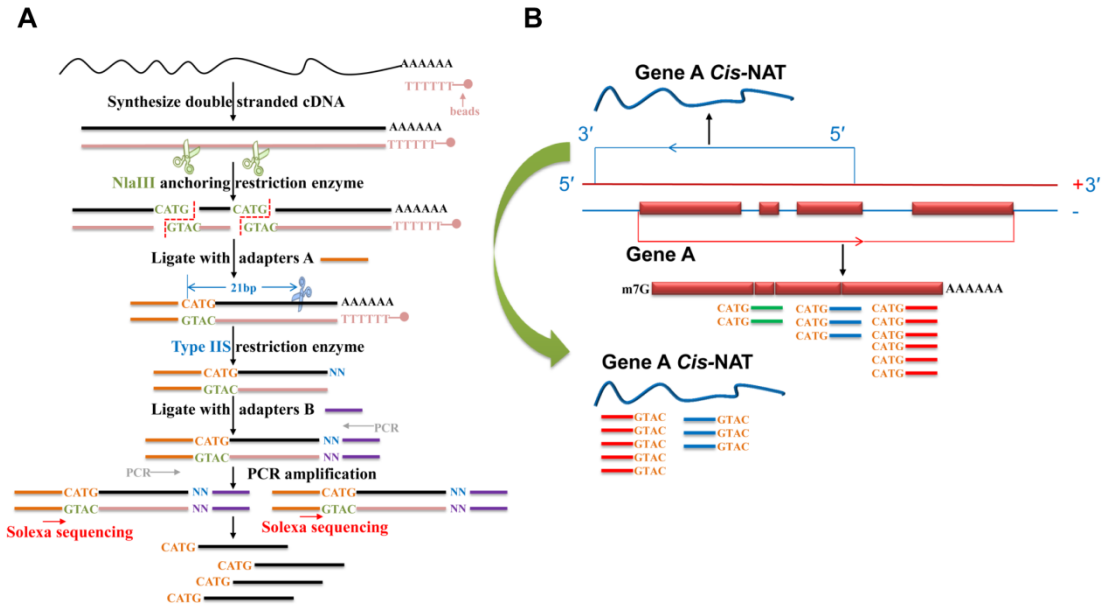

Figure S1. Outline of DGE library construction and 21 bp tag mapping back to the original mRNA.

(A) Outline of library construction. Each mRNA (black and curved line) is captured by oligo(dT) beads to undergo double-stranded cDNA synthesis. NlaIII (green scissors) is used to digest the cDNA, and a 4 bp overhang (GTAC) remains. Adapter A (brown) is ligated to the overhang of oligo(dT) bead-anchored cDNA fragments and adds a recognition site for the TypeIIS tagging enzyme MmeI. After MmeI (blue scissor) digestion, adapter B (purple) is ligated to the resulting 2 bp overhang. The PCR primers (grey arrow) that anneal to adapters A and B are utilized to enrich tags. The last two steps include cluster generation and sequencing (red arrow). (B) 21 bp tag mapping back to the original mRNA. NlaIII is used in DGE library construction to cleave a double-stranded cDNA molecule with 5′-CATG-3′ into fragments. Only the last fragment linked to oligo(dT) beads will be ligated to the adapter and cleaved by MmeI from the 3′-most NlaIII site to generate a 21 bp tag. However, the cleavage efficiency of NlaIII is not 100% at each site, which indicates that some known genes and *cis*-NATs contain more than one 21 bp tag. Nonetheless, the general trend is that the 3′ CATG site of RNA is more likely to be cleaved. Therefore, the TPM value of the 5′ terminal-located tags becomes lower than that the 3′ terminal-located tags.

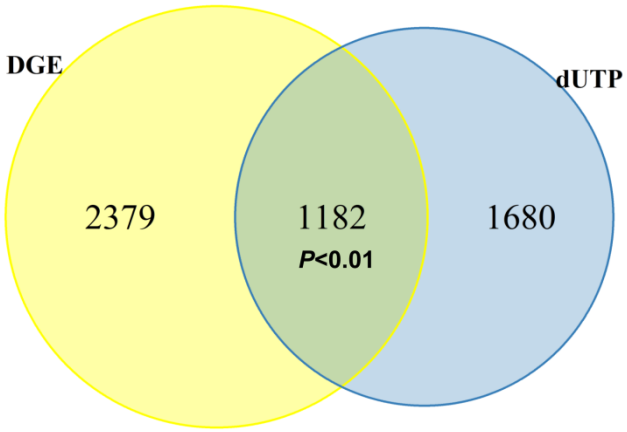

Figure S2. Representation of SA genes identified from DGE and dUTP libraries.

The intersection of SA genes between the DGE and dUTP libraries was tested by adopting Fisher's exact test.

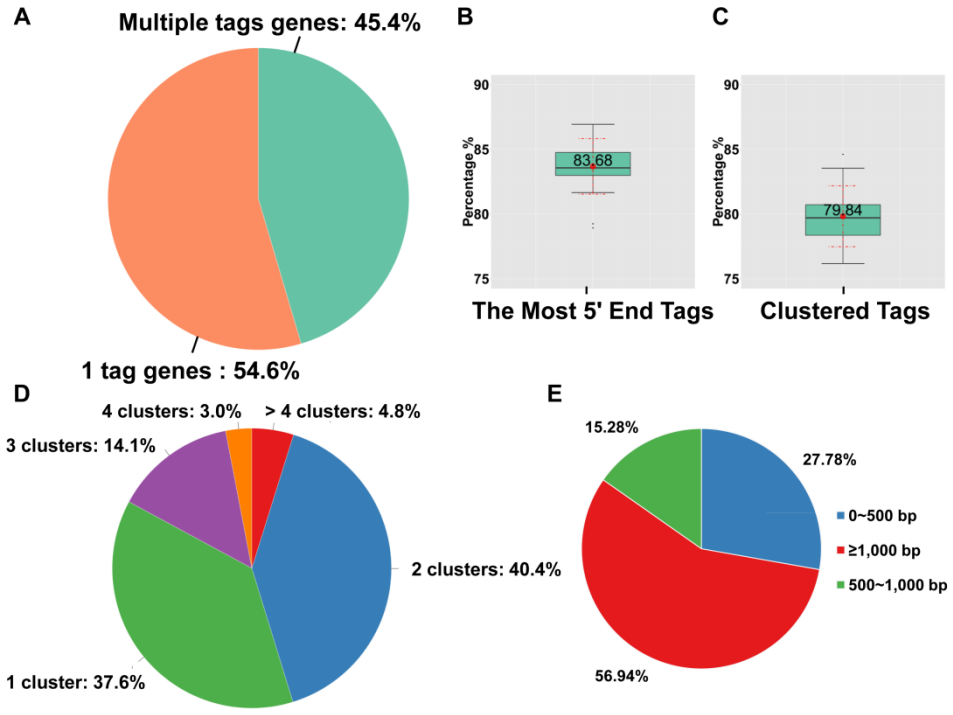

Figure S3. Statistical results of expression characteristics of gene sense strand mapped tags.

(A) Proportion of genes containing different counts of mapped tags. We found that 54.6% of the genes contained only one mapped tag, which was defined as multiple tag genes. (B) Boxplot of expression pattern of the mapped tags of multiple tag genes. The average percentage in the boxplot among 20 libraries was calculated by the most RNA 5′ end-located tags whose expression levels were lower than other mapped tags of the same RNA. (C) Boxplot of the average percentage of clustered tags. The tags of each multiple tag gene were sorted by their mapping position from 5′ to 3′ in the transcript of the gene. Neighboring tags were clustered for their TPM of 5′ tag being less than their 3′ tag, and these neighboring tags were defined as clustered tags. (D) Statistics of gene clusters. Each gene could contain one or more clusters. The most enriched clusters were two. (E) Statistical results of length distribution of the *cis*-NATs. Statistical tests were performed by using 864 *cis*-NATs, which were identified using both DGE and dUTP data.

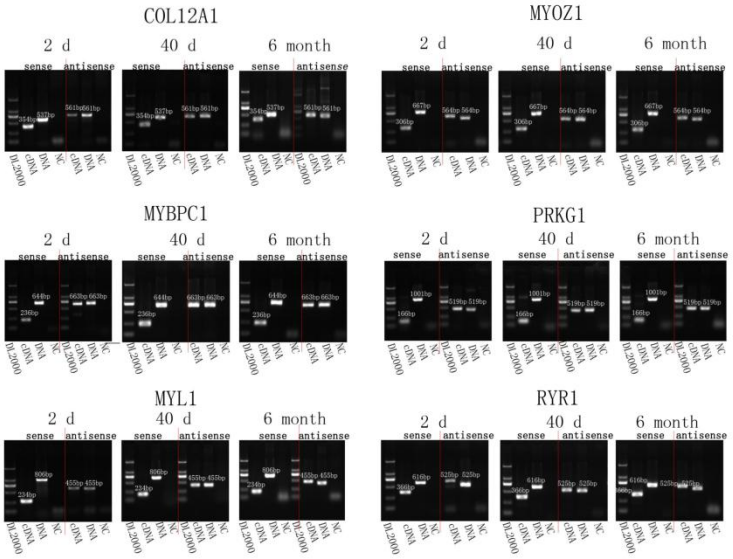

Figure S4. RT-PCR verification of *cis*-NATs and their sense genes.

The transcript regions of the six *cis*-NATs all contained part of the intron region of their sense genes. At least one primer of the six *cis*-NATs was designed to bind to the intron region of their sense genes to avoid the amplification of their sense genes. For the amplification of sense transcription, one pair of primers was designed to amplify both the cDNA and DNA templates. The PCR results of all the six genes showed that the cDNA was not contaminated with DNA for a product from the cDNA that is smaller than the DNA template.

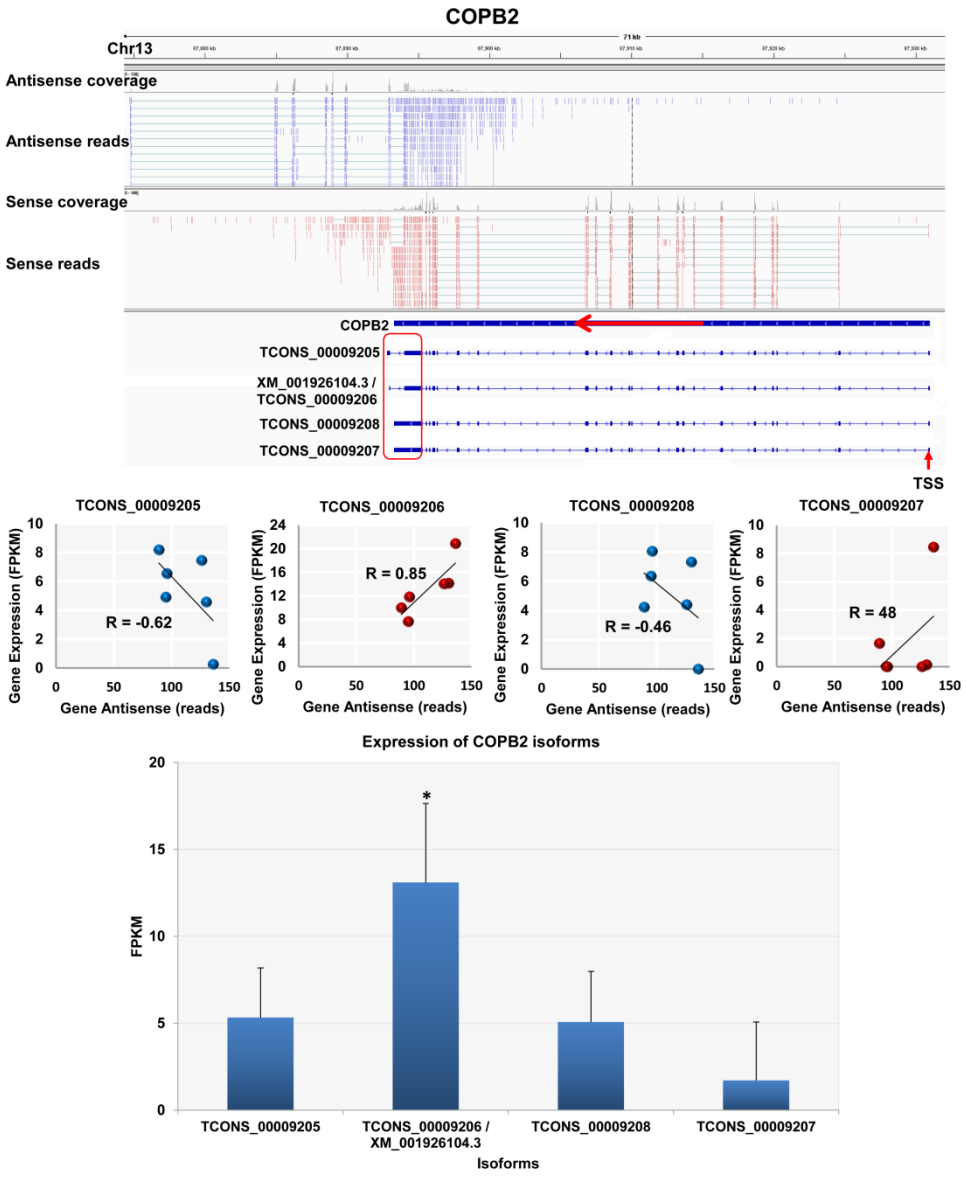

Figure S5. Illustration of *cis*-NATs affecting the alternative splicing of their sense genes.

FPKM means fragments per kilobase of exon per million fragments mapped. TSS means transcription start site. Student's t-test was employed to detect the expression difference between the various isoforms of the *COBP2* gene. The expression of the XM_001926104.3 or TCONS_00009206 isoform was significantly higher than that of the other isoforms of the *COBP2* gene.

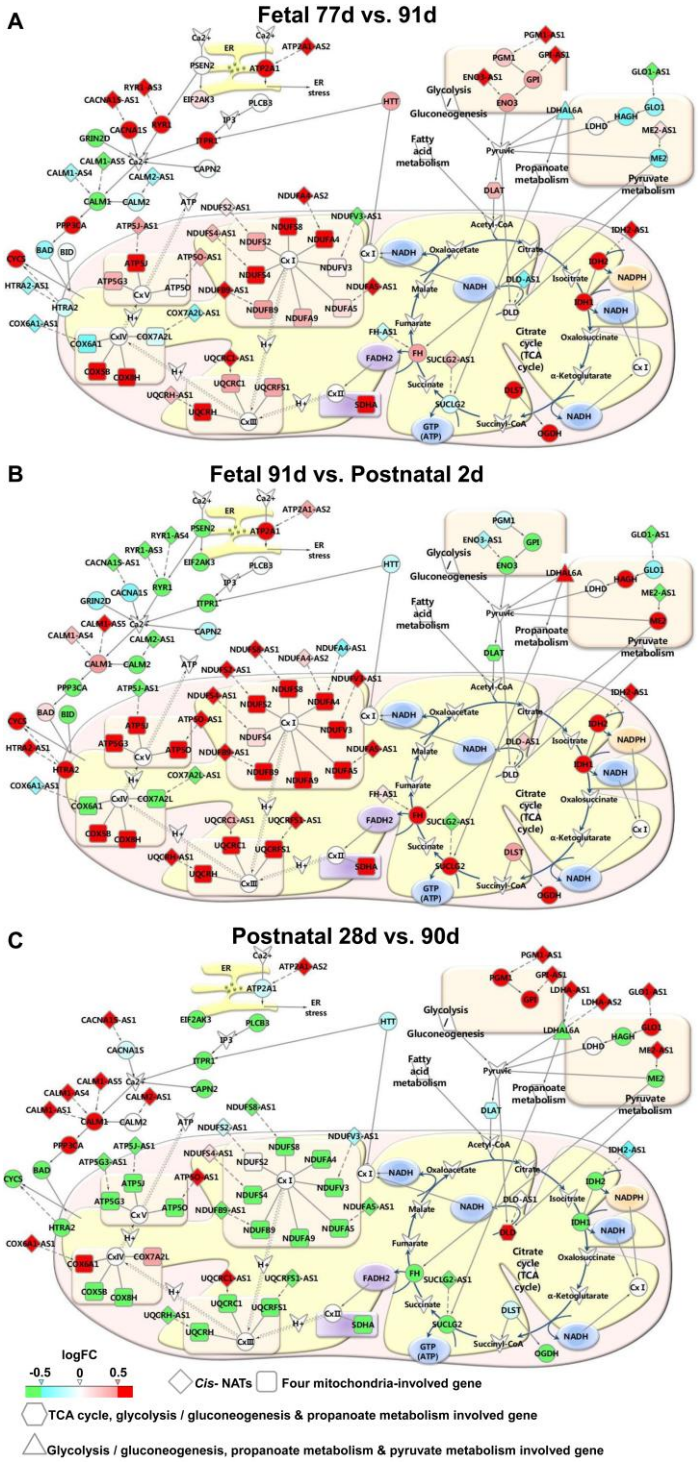

Figure S6. Signal-flow of energy metabolism pathway-involved genes of LT pigs during different muscle development stages.

Green, blue, pink and red fill color represents $\log_2FC \leq -0.5$, $-0.5 < \log_2FC \leq 0$, $0 < \log_2FC < 0.5$, and $\log_2FC \geq 0.5$, respectively. LT, Lantang. (A-B) Signal-flow analyses of energy metabolic pathways-involved genes from fetal day 77 to postnatal day 2. (C) Signal-flow analyses of e energy metabolic pathways-involved genes from postnatal day 28 to day 90.

Supplementary Data

Supplementary Data 1

File name: Supplementary Data 1.xls
Description of data: gff file of cis-NATs
File format: .xls (This file can open with EXCEL and TXT editor)

Supplementary Data 2

File name: Supplementary Data 2.xls
Description of data: Confirmation of the cis-NATs by using NCBI least annotated data
File format: .xls

Supplementary Data 3

File name: Supplementary Data 3.xls
Description of data: Correlated cis-NATs and genes pairs
File format: .xls

Supplementary Data 4

File name: Supplementary Data 4.xls
Description of data: Correlated cis-NAT and isoform pairs of SA genes
File format: .xls

Supplementary Data 5

File name: Supplementary Data 5.xls
Description of data: Significant different expressed transcripts
File format: .xls

Supplementary Data 6

File name: Supplementary Data 6.xls
Description of data: Significant enrichment of BP GO terms of correlated cis-NATs of Landrace and Lantang pigs
File format: .xls

Supplementary Data 7

File name: Animate signal-flow of energy metabolism pathway-involved genes during different muscle development stages.
File format: .WMV (This file can open with Windows Media Player (on windows) or MPlayer OSX (on Mac))
Description of data:
Green, blue, pink and red fill color of notes represented log2FC $\leq$ -0.5, -0.5 < log2FC $\leq$ 0, 0 < log2FC < 0.5 and log2FC $\geq$ 0.5, respectively. LR, Landrace; LT, Lantang.