Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Identification and validation of differentially expressed genes from pig skeletal muscle

L.L. Verardo[1], C.S. Nascimento[1], F.F. Silva[2], E. Gasparino[3], M.F.Martins[4], E. Toriyama[1], V.R. Faria[1], M.E. Botelho[1], K.A. Costa[1], P.S. Lopes[1] & S.E.F. Guimarães[1]

1 Department of Animal Science, Universidade Federal de Viçosa – UFV, Viçosa, MG, Brazil
2 Department of Statistics, Universidade Federal de Viçosa – UFV, Viçosa, MG, Brazil
3 Department of Animal Science, Universidade Estadual de Maringá – UEM, Maringa, PR, Brazil
4 EMBRAPA Gado de Leite, Juiz de Fora, MG, Brazil

Keywords
Myogenesis; prenatal life; real-time PCR; skeletal muscle.

Correspondence
S.E.F. Guimarães, Departamento de Zootecnia, Universidade Federal de Viçosa, av. PH Rolfs, s/n, 36570-000, Viçosa, Brazil. Tel: (+55) 313138992273; Fax: (+55) 313138992275; E-mail: sfacioni@ufv.br

Received: 27 February 2012;
accepted: 26 July 2012

Summary

Pig is an important animal for meat production; this is generally associated with characteristics determined prenatally during myogenesis. Expressed sequence tags (EST) can provide direct information on the transcriptome and indirect information on the relation between the genome and phenotype, giving information about differentially expressed genes (DEG). In this work, the identification and annotation of DEG from EST libraries of three pig breeds (Duroc, Large White and Local Breed Piau) were performed followed by real-time PCR analyses during pre- and postnatal stages (21, 40, 70 and 90 days of pregnancy and 107, 121 and 171 days postnatal) from commercial breed animals for analysis of genes expression levels. Therefore, 34 genes differentially expressed were identified, of which 21 grouped in a network related with muscle development. From this, the expression profile of 13 genes was measured, to confirm their relationship with myogenesis like ANKRD2, MYBPC1, NEB and MYL2. These genes showed a prenatal high expression in this study. Besides, novels candidates for muscle development (TP53 and DCTN1) were listed. These findings can contribute to better explaining gene function mechanism and are helpful in uncovering the pathways that mediate pre- and postnatal skeletal muscle development in vertebrates.

Introduction

Pig (*Sus scrofa*) is an important animal for meat production being the most widely consumed meat in many countries (Foreign Agricultural Service/USDA Office of Global Analysis, 2011). Studies in foetal pig skeletal muscle have revealed developmental patterns of gene expression, including genes not previously associated with myogenesis (Sollero *et al.* 2011). In this way, more detailed studies of this process seeking for a better meat production still have a particular place in the field of animal genetics research.

The interest in muscle growth potential is generally associated with characteristics determined prenatally during myogenesis (Rehfeldt *et al.* 2000). This lean muscle growth potential of an animal largely depends on the number of muscle fibres prenatally formed, being the postnatal increase in muscle fibre size limited by genetic and physiological reasons (Rehfeldt *et al.* 2004). One of the unique characteristics of the skeletal muscle is the diversity on its morphological and biochemical properties (Ryu *et al.* 2006), which may be explained by the type of protein present according to the expressed genes. The genes expression can be evaluated by

doi:10.1111/jbg.12006

analyses of expressed sequence tags (EST) present in specific populations.

In different areas of research, ESTs and full-length cDNAs provide direct information on the transcriptome and indirect information on the relation between the genome and different phenotypes (Gorodkin *et al.* 2007). This resource can be used in a wide range of applications (Murray *et al.* 2007), for example, to identify genes of importance in meat production or to analyse the effect of a specific gene in the muscle development. Nevertheless, the power of these libraries as a comprehensive and quantitative transcript profiling method relies on efficient computational tools for data generation, management and analysis, plus a good-quality experimental validation of these genes through methods such as the real-time PCR.

The aim of the present study was using our previous results from an *in silico* identification of differentially expressed genes (DEGs) in EST libraries of three different pig breeds, Duroc, Large White and Local Breed Piau, described in Nascimento *et al.* (2012) to better understand the dynamics of gene networks during muscle development in the pig, comprising not only prenatal but also postnatal stages. The gene expression profile was analysed through real-time PCR in white commercial breed animals.

Material and methods

All methods involving animal handling were carried out in accordance with regulations approved by the institutional animal welfare and ethics/protection commission of the Federal University of Viçosa (UFV; DVT-UFV 02/2008).

EST dataset

The ESTs used in this study were generated from three different pig breeds' skeletal muscle tissue cDNA libraries (Duroc, Large White and a Local Breed Piau) constructed by Nascimento *et al.* (2012).

Differentially expressed genes identification

First, ESTs were submitted to CAP3 program (http://deepc2.psi.iastate.edu/aat/cap/cap.html), for clustering and assembling into contigs. Functional annotation of ESTs for the three datasets was performed using the program BlastX (Altschul *et al.* 1997), against the data base of transcripts from SwissProt (http://www.expasy.ch/sprot/). More details can be obtained at Nascimento *et al.* (2012).

The results for each gene identified were normalized using the total number of ESTs in each library and defined as values of relative expression for the identification of DEGs using the IDEG6 program (Romualdi *et al.* 2003), available online at the web server (http://telethon.bio.unipd.it/bioinfo/). This software calculates the values of six different statistical tests used for the identification of DEGs in multiple tag sampling experiments. For pairwise comparisons, we applied the AC statistic (Audic and Claverie, 1997), Fisher's 2 × 2 exact test and 2 × 2 chi-square test, and in multiple comparisons the R statistic (Stekel *et al.*, 2000), GT statistic (Greller and Tobin, 1999) and general chi-square test were used. For each test, the associated significance thresholds ($p < 0.05$) were applied using the Bonferroni correction and the DEGs were identified, once they were significantly expressed in at least one test.

Gene network analysis

Aiming to examine the process of shared pathways, the Ensembl gene identifiers were used. These gene identifiers were extracted from Ensembl Biomart web site (http://www.ensembl.org/biomart/martview/02295617921911c90441a7a165cd0f7f). The program toppcluster (http://toppcluster.cchmc.org/) was used to obtain the functional Gene Ontology (GO), identifying the biological mechanisms and pathways and functions involving the DEGs. The application Cytoscape (www.cytoscape.org/) was used to visualize and edit the identified pathways. Only networks that included direct relationships between genes and muscularity have been maintained. Genes, from the EST libraries, were then selected for validation by real-time PCR based on their differential expression.

Sampling

A total of 21 *Longissimus dorsi* (LD) samples divided in pre- and postnatal stages were collected. LD was used, as it has a high importance in Brazilian pork market and adequate samples were available at the time of this trial. At the UFV Pig Breeding Farm, pregnant commercial gilts at 21, 40, 70 and 90 days of gestation were aborted using the protocol described at Sollero *et al.* (2011). Briefly, intramuscular injections of 1 ml Prelobam® (Intervet, São Paulo, Brasil) (PGAF-α)-plus 1 ml Estrogen were administered followed by by 2 ml Orastina® (Intervet) (Ocitocine) 12 h later. LD muscle samples were isolated from 12 foetuses, 21d (n = 3), 40d (n = 3), 70d (n = 3) and 90d (n = 3), and placed in sterile tubes containing RNA*later*® (Ambion,

Carlsbad, CA, USA). Samples were stored at 4°C overnight and at −70°C prior to RNA isolation. In the same way, LD samples from commercial castrated males in three different ages, 107 days (n = 3), 121 days (n = 3) and 171 days (n = 3) postnatal, were isolated and stored.

RNA isolation and reverse transcription

Total RNA was isolated from the muscle with RNeasy® Mini Kit (Qiagen, Valencia, CA, USA). The total concentration of RNA was estimated in a spectrophotometer NanoVueTM Plus (GE Healthcare, Freiburg, Germany) and quality at the Agilent 2100 Bioanalyzer© (Agilent Technologies, Palo Alto, CA, USA.) obtained an 7.6 average RIM quality value. Details about RNA extraction were described by Serão *et al.* (2010) and Sollero *et al.* (2011). The first strand of cDNA synthesis was performed using ProtoScript® M-MuLV First-Strand cDNA Synthesis Kit (New England Biolabs Inc., Beverly, MA, USA), and its concentration was estimated in a spectrophotometer NanoVueTM Plus (GE Healthcare).

Quantitative real-time PCR

The real-time PCR was performed in thermal cycler ABI Prism 7300 Sequence Detection Systems (Applied Biosystems, Foster City, CA, USA) using GoTaq® qPCR Master Mix (Promega Corporation, Madison, WI, USA). Gene target sequence was recovered from nucleotide sequences obtained from the GenBank database (www.ncbi.nlm.nih.gov) and used to construct primers by the PRIMERQUEST program available at (www.idtdna.com/SciTools/Applications/primerQuest) provided by Integrated DNA Technologies, Inc (Coralville, IA, USA). Nucleotide sequence of the primers, accession numbers and the amplicon length are summarized in supplementary data (Table S1).

Amplification conditions for all systems were as follows: 95°C for 2 min, 40 cycles of denaturation at 95°C for 15 s and annealing/extension at 60°C for 60 s. The efficiencies of amplification represented around 100% in each cycle, and the relative abundance was calculated using an equation to correct differences in efficiency as described by Pfaffl 2001. Concentrations of primer and cDNA are summarized in supplementary data (Table S2). Housekeeping genes such as those encoding β-actin (*BACT*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), commonly used as internal controls for such analysis, were not suitable for normalization in the experiment because their transcription is altered during myogenesis (Radoniæ *et al.* 2004). Titin (TTN), which was consistently expressed in our study according to GENORM program (Vandesompele *et al.* 2002), was therefore used as an internal control.

The amplification of target genes was performed at different wells and in duplicates. The used experimental design was completely randomized, with three samples per period (21, 40, 70 and 90 days of pregnancy; 107, 121 and 171 days after birth of castrated white commercial males). Initially, data from real-time PCR were analysed using the linear mixed model, described below.

$$y_{ijk} = GP_{ik} + D_{ijk} + e_{ijk}$$

Where y_{ijk} is the measured expression level of gene i on animal j in the period k; GP_{ik} is the effect of gene i in period k; D_{ijk} is a random sample-specific effect (common to both genes), $D_{ijk} \sim N(0, \sigma_D^2)$; and e_{ijk} is a residual term, $e_{ijk} \sim N(0, \sigma_e^2)$.

All statistical procedures were performed using SAS 9.0 for Windows (Statistical Analysis System Institute, Inc., Carry, NC, USA). The routine QPCR_MIXED: SAS® developed to generate commands in SAS PROC MIXED was used to analyse real-time PCR data in the analysis (Steibel *et al.* 2009). For each target gene, the comparison of expression values from a period to the following was performed by CONTRAST statement of the GLM procedure using Student's *t*-test.

Clustering relative quantification of gene expression

Hierarchical clustering was performed on real-time PCR data. ΔC_t values (target C_t – internal control C_t) were used in the analysis. The set of muscle-specific genes expression values across seven periods was used for clustering based on complete linkage method with Pearson correlation as a distance (Eisen *et al.* 1998).

Results

Differentially expressed genes identification

The CAP3 sequence assembly program was used to group redundant ESTs, which had overlapping sequences from all three libraries. A consensus sequence was obtained for each contig, and every EST present in the contig was considered a copy of the transcript from the same gene sequence. A number of 3670 unique sequences representing putative transcripts from pig breeds forming 905 contigs (merged

overlapping sequences) and 2765 singletons were obtained, defined as sequences that did not assemble into contigs using the defined assembly parameters. More details about cDNA Library construction, reads sequencing and sequence assembly can be obtained at Nascimento *et al.* (2012).

After BlastX of all 905 contigs against SwissProt data base, it was performed IDEG6 analysis. Comparison of the three sequenced cDNA libraries from divergent genetic background identified sequences that were differentially expressed among them. A total of 54 DEGs were identified, 34 representing coding genes for known proteins (Table 1), and were chosen for the subsequent analysis described in the present paper.

Gene networks analyses

To understand the functions of the 34 selected DEGs, we collected information about their biological process, cellular component and molecular function in the Gene Ontology. Furthermore, using the application TOPPCLUSTER, we were able to identify metabolic pathways and interaction. Thus, it was possible to identify genes with obvious roles in muscle physiology allowing inserting these genes into relevant functional metabolic networks with the addition of five other related ones. Among the 34 genes analysed, 16 genes (MYH2, DCTN1, PDLIM5, TPM2, MYBPC1, TNNT3, ACTA1, MYL2, MYH1, MYH7, SMARCD3, ANKRD2, MYLPF, ANK1, TRIM54 and MYBPH) and

Table 1 Description of the 34 transcripts with functional annotation that showed differential expression tested by IDEG6 program

Library[a]	Gene	Tests[b]
Large White	MT-CYB	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, GT, R
	MyH1	AC1–2, AC1–3, AC2–3, Fisher1–3, Fisher2–3, Chi^{2}2–3, R, Chi^2
	MyH2	AC1–2, AC1–3, Fisher2–3, Chi^{2}2–3, R, Chi^2
Piau	ANK1	AC1–2, AC1–3, AC2–3, Fisher1–3, Fisher2–3, Chi^{2}1–3, Chi^{2}2–3, R, Chi^2
	MyLPF	AC1–3, AC2–3, Fisher1–3, Fisher2–3, Chi^{2}1–3, Chi^{2}2–3, R, Chi^2
	NDUFB2	AC2–3, R
	TNNT3	AC1–2, AC1–3, AC2–3, Fisher1–3, Fisher2–3, Chi^{2}1–3, Chi^{2}2–3, R, Chi^2
	TPT1	AC1–3, AC2–3, Fisher1–3, Fisher2–3, Chi^{2}1–3, Chi^{2}2–3, R, Chi^2
Duroc	ACTA1	AC1–3
	ACY1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	ANKRD2	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	ATP5A1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	ATP6V1H	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	CA3	AC1–3
	CGGBP1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	DCTN1	AC1–2, AC1–3, AC2–3, Fisher1–2, Fisher2–3, Chi^{2}1–2, Chi^{2}2–3, R, Chi^2
	DYSFIP1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	ENO3	AC1–3, AC2–3, Fisher1–2, Fisher1–3, Chi21–3
	EXOSC8	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	MT-COX3	AC1–3, AC2–3, Fisher1–2, Fisher1–3, Chi^{2}1–3
	MyBPC1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	MyBPH	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	MyH7	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	MyL2	AC1–3
	NAGA	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	NDUFA9	AC1–2, AC1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	PDLIM5	AC1–2
	RPL10A	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	SDHS	AC1–2, AC2–3, Chi^{2}1–2
	MARCD3	AC1–2, AC1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	TCBE2	AC1–2, AC1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	TPM2	AC1–2, AC1–3
	TRIM54	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2
	ZCRB1	AC1–2, AC1–3, Fisher1–2, Fisher1–3, Chi^{2}1–2, Chi^{2}1–3, R, Chi^2

[a]According to Nascimento *et al.* (2012).

[b]IDEG6 tests that showed differential expression, AC: Audic and Claverie statistic, Fisher: Fisher test, GT: Greller and Tobin statistic, R: R statistic and Chi^2: chi-squared test. Tests with 1–2, 1–3 and 2–3 means pairwise comparisons between Large White – Piau, Large White – Duroc and Piau – Duroc libraries, respectively. GT, R and chi-square tests were used for multiple comparisons.

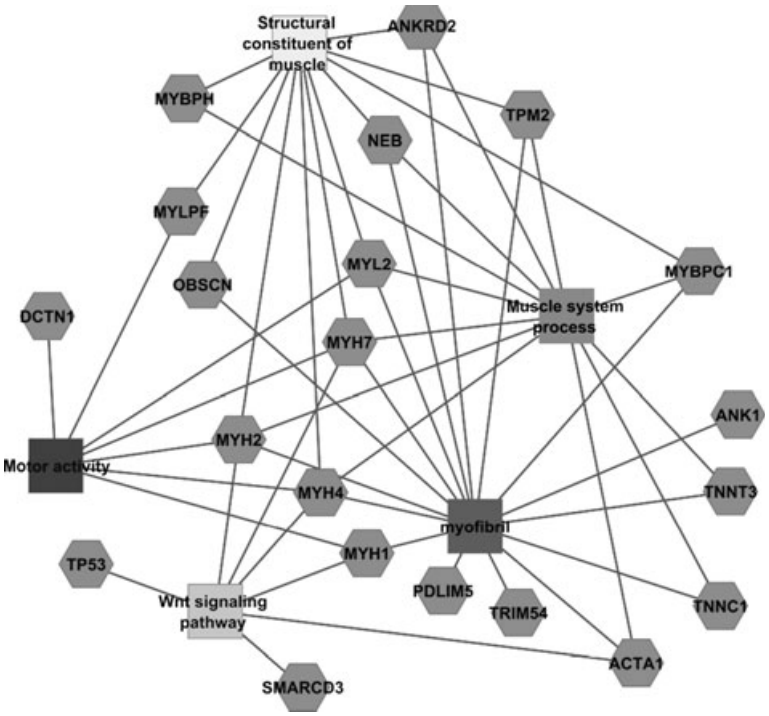

Figure 1 Functional gene networks and their interactions. It describes the relationships between 16 differentially expressed genes and other five included (⬣), there are five important muscle development–related subnets; motor activity (■), structural constituent of muscle (□), muscle system process (■), myofibril (■) and Wnt signalling pathway (□).

other five genes added owing to their presence at the same metabolic path (MYH4, TP53, TNNC1, NEB and OBSCN) could be grouped into networks of functional relevance for muscularity (Figure 1). These are myofibril, muscle system process, structural constituent of muscle, motor activity and Wnt signalling pathway. The function of the 16 DEGs and the other five added genes is described through the Table 2 using the assignment of GO terms.

Quantitative real-time PCR of DEGs

From the 21 genes selected above, real-time PCR was consistently performed across seven periods (four prenatal and three postnatal) for 13 genes (Table S1). The pattern of expression for each gene is reported by hierarchical clustering describing the expression level gene by gene to better evaluate differences between periods for each specific gene during embryo growth and postnatal life. During prenatal stage, nine genes showed a higher expression ($p < 0.05$) in at least one period (ANKRD2, MYBPC1, NEB, DCTN1, MYL2, TP53, TPM2, OBSCN and ANK1) and four during postnatal (MYH2, ACTA1, MYH7 and TNNT3). Unsupervised hierarchical clustering based on the expression values (ΔC_t) across seven periods (21, 40, 70 and 90 days prenatal and 107, 121 and 171 days of adult age 1–7, respectively) have grouped all genes into four separate clusters. One containing only postnatal expression and the other three with a variable expression profile having at least one high expression during prenatal (displayed as the heat map in Figure 2). A probability value for each contrast (gene/period versus period) demonstrates the significance of each contrast ($p < 0.05$); these data are showed in Table 3.

Discussion

The IDEG6 analysis showed that the chi-square test was the most sensitive, recovering most of the significant cases. In contrast, the GT test alone was not sufficient to detect subtle divergences in the analysed data. This observation was consistent with results based on theoretical and observed data from Romualdi *et al.* 2001, who considered the use of AC and chi-square as the most appropriate combination to test differential distribution in multiple tag sampling experiments with cDNA libraries.

At the network proposed here, sixteen proteins are connected to the core network with the subnet myofibril being the subnet with more proteins (Figure 1). The five remaining proteins are connected to the others by the subnets muscle system process, structural constituent of muscle, motor activity and Wnt signalling pathway. The latter subnet is a network of proteins best known for their roles in embryogenesis and cancer, but also involved

Table 2 Pathway, biological process, molecular function and cellular component from genes represented in the network

Genes	Pathway	GO: Biological Process	GO: Molecular Function	GO: Cellular Component
MYL2	Genes involved in striated muscle contraction	Muscle system process	Structural constituent of muscle/ motor activity	Myofibril
MYBPC1	Genes involved in striated muscle contraction	Muscle system process	Structural constituent of muscle	Myofibril
ACTA1	Wnt signalling pathway	Muscle system process	Structural molecule activity	Myofibril
MYH2	Wnt signalling pathway	Muscle system process	Structural constituent of muscle/ motor activity	Myofibril
MYH7	Wnt signalling pathway	Muscle system process	Structural constituent of muscle/ motor activity	Myofibril
TNNT3	Genes involved in striated muscle contraction	Muscle system process	Cytoskeletal protein binding	Myofibril
TPM2	Genes involved in muscle contraction	Muscle system process	Structural constituent of muscle	Myofibril
TP53	Wnt signalling pathway/Huntington's disease	Protein import into nucleus, translocation	DNA strand annealing activity	Chromatin
DCTN1	Vasopressin-regulated water reabsorption	G2/M transition of mitotic cell cycle	Motor activity	Cytoskeletal part
NEB	Genes involved in muscle contraction	Muscle system process	Structural constituent of muscle	Myofibril
OBSCN	–	Carbohydrate metabolic process	Structural constituent of muscle	Myofibril
ANK1	–	Exocytosis	Cytoskeletal protein binding	Myofibril
ANKRD2	–	Muscle system process	Structural constituent of muscle	Myofibril
MYH4	Wnt signalling pathway	Muscle system process	Structural constituent of muscle/ motor activity	Myofibril
PDLIM5	–	Regulation of synaptogenesis	Cytoskeletal protein binding	Myofibril
MYH1	Wnt signalling pathway	–	Motor activity	Myofibril
TNNC1	Genes involved in striated muscle contraction	Muscle system process	Cytoskeletal protein binding	Myofibril
SMARCD3	Wnt signalling pathway	Muscle structure development	Transcription coactivator activity	Nucleus
MYLPF	Tight junction	Muscle structure development	Structural constituent of muscle/ motor activity	Myosin complex
TRIM54	–	Microtubule-based process	Cytoskeletal protein binding	Myofibril
MYBPH	–	Muscle system process	Structural constituent of muscle	Myosin complex

in normal physiological processes in adult animals (Lie *et al.*, 2005).

Prenatal over-expressed genes

During embryo growth, there is the occurrence of two muscle development waves, the first during 35–55 days and the other during 55–90 days of pregnancy (Wigmore & Stickland 1983). Here, two genes (MYL2 and MYBPC1) had high expression in critical periods of prenatal stage (day 40 and days 70 and 90, respectively) forming the Cluster 1; at the network described above, they are connected with myofibril, muscle system process and structural constituent of the muscle as expected with the exception of MYL2, which is also related to motor activity.

The MYL2 gene expression profile in this study contrasting with 21d/40d and 40d/70d showed a single peak of high expression at 40 days of pregnancy (Figure 2), which coincides with the primary generation stage of muscle formation which is between 35 and 60 days of pregnancy (Wigmore & Stickland 1983). Our findings are in agreement with Zhang *et al.* 2009, supporting the hypothesis of MYL2 to be involved in myogenesis.

The MYBPC1 gene expression profile showed a low expression at the first period (21d/40d) getting higher until 70 days of pregnancy, when it started to be low again. This period from 70 to 90 days coincides with the second generation of muscle fibre formation, which is between 54 and 90 days of pregnancy (Wigmore & Stickland 1983).

Three genes (TPM2, TP53 and DCTN1) showed a very low expression during postnatal stage (Cluster 3 on Figure 2). The TPM2 gene is showed to be differentially expressed across the stages, which is more

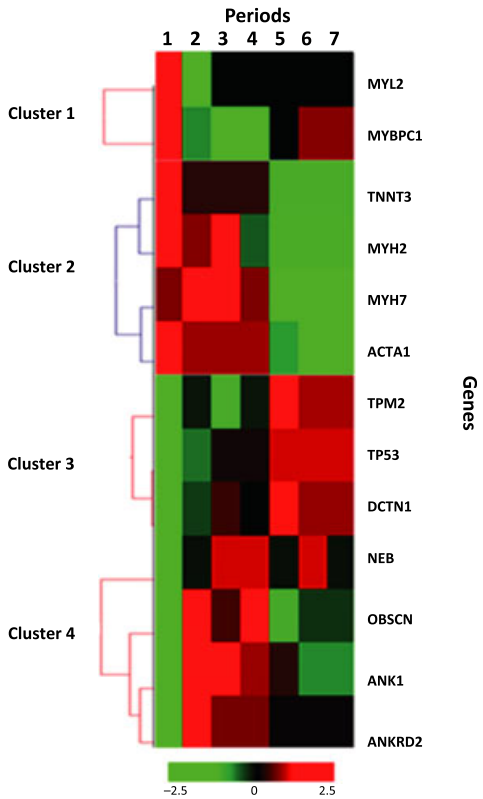

Figure 2 Unsupervised hierarchical clustering of muscle differentially expressed genes expression levels across all seven periods (1–4 prenatal, 21, 40, 70 and 90 days of pregnancy, respectively, and 5–7 postnatal, 107, 121 and 171 days after birth). They were divided in four groups by similar expression pattern. A median expression value equals to zero was designated in black; green increased expression and red reduced expression. The full colour figure can be found online at http://onlinelibrary.wiley.com.

expressed during prenatal, demonstrating that it is playing a role during muscle development. The postnatal expression profile found in this work agrees with the results found by Różycki *et al.* 2011, where the expression level of TPM2 did not change significantly in Polish commercial line animals at slaughter age.

The TP53 gene is connected only with the subnet Wnt signalling pathway, which is strongly related with embryogenesis path according to its expression profile. This gene showed to be highly expressed at the first period (21-day prenatal) going down at 40 days prenatal, and differing from all three postnatal periods. The DCTN1 had a high expression at the first period (21d) but its expression at 40 days was also high when compared with the other five periods (Table 2). This high expression level during the first and second period seems to play a role during the embryo stage.

According to gene expression profile, it was possible to see a cluster (Cluster 4) with four genes (ANKRD2, NEB, OBSCN and ANK1) having a very high expression level at 21 days of gestation. In Figure 2, we can see the ANKRD2 and NEB expression level being high at 21 days and after that being low across all other six periods. Differing on this cluster we have two genes, OBSCN and ANK1. Both gene expression profiles had a higher point at 21 days and lower point during the rest of the embryo stage and getting higher again during postnatal stages, with the exception of the fifth period (107 days of adult age) in which ANK1 had almost the same expression level as period four (90 days of pregnancy). There are reports about OBSCN gene where it is positively correlated with intra muscular fat content in adult pigs (Serão *et al.* 2010). On our network, it was connected with myofibril and structural constituent of the muscle subnets.

Postnatal over-expressed genes

On Cluster 2, four genes were found (MYH2, MYH7, ACTA1 and TNNT3), which may play a role in hypertrophy, once a higher expression during postnatal stage is observed for them. Here, MYH2 and MYH7 are at the same subnets, also with Wnt signalling pathway even their expression level showing to be high only during postnatal stage (Figure 2). The MYH2 gene had two points during prenatal stage being significantly less expressed, which was at 21d and 70d, while MYH7 gene had all prenatal periods being less expressed than postnatal.

As well as MYH2 and MYH7, ACTA1 and TNNT3 genes showed a similar expression profile. ACTA1 was linked to the Wnt signalling pathway, myofibril and muscle system process. Here, its expression profile showed low levels at prenatal stage and high at postnatal. There are reports where the expression levels of TNNT3 gene showed to be not differentially related to the pig slaughter age of Polish commercial line animals (Różycki *et al.* 2011). In our findings, this gene had a higher expression during postnatal stage while the expression level in prenatal stage was lower, which means that this gene even though does not differ during postnatal periods may play an important role on that stage.

Only few genes analysed at the present study have been previously described in relation to their expression profile during the mentioned stages in skeletal muscle. We could find a few studies during postnatal life, which agree with our results about TPM2 and TNNT3 genes (Różycki *et al.* 2011). On the network designed here, these two genes are connected by

Table 3 p Values for differential expression values for each gene in the comparison between two periods

	Differentially expressed genes[a]												
Contrasts[b]	ANKRD2	MYH2	ACTA1	DCTN1	MYBPC1	NEB	TPM2	OBSCN	ANK1	MYH7	MYL2	TNNT3	TP53
21d/40d	*0.0028*	0.2354	0.1414	*<0.0001*	*0.0027*	*0.0022*	0.0682	*<0.0001*	*<0.0001*	0.2680	*0.0010*	*0.0002*	*<0.0001*
21d/70d	*0.0099*	0.6030	*0.0061*	*<0.0001*	*0.0005*	*0.0002*	0.1285	*<0.0001*	*<0.0001*	0.1841	0.1195	*<0.0001*	*<0.0001*
21d/90d	*0.0151*	*0.0438*	*0.0041*	*<0.0001*	*0.0007*	*0.0003*	0.0682	*<0.0001*	*<0.0001*	0.3284	0.0840	*<0.0001*	*<0.0001*
21d/107d	*0.0366*	*0.0096*	*<0.0001*	*<0.0001*	*0.0103*	*0.0023*	*<0.0001*	0.1355	*<0.0001*	*0.0065*	0.0625	*<0.0001*	*<0.0001*
21d/121d	*0.0270*	*0.0123*	*<0.0001*	*<0.0001*	0.0813	*0.0004*	*0.0002*	*0.0275*	*0.0013*	*0.0048*	0.1133	*<0.0001*	*<0.0001*
21d/171d	*0.0222*	*0.0232*	*<0.0001*	*<0.0001*	0.0824	*0.0028*	*<0.0001*	*0.0303*	*0.0005*	*0.0091*	0.4237	*<0.0001*	*<0.0001*
40d/70d	0.6101	0.4907	0.1180	*0.0190*	0.4101	0.2258	0.7243	0.0512	0.7613	0.8111	*0.0252*	0.0647	0.0702
40d/90d	0.4929	0.3459	0.0835	*0.0286*	0.4883	0.3269	0.9995	0.0945	0.1415	0.8900	*0.0371*	0.2500	*0.0297*
40d/107d	0.2829	0.1009	*0.0001*	*<0.0001*	0.5160	0.9759	*0.0014*	*<0.0001*	*0.0137*	*0.0007*	0.0505	*0.0003*	*0.0002*
40d/121d	0.3480	0.1250	*<0.0001*	*0.0005*	0.1021	0.3842	*0.0077*	*0.0001*	*0.0001*	*0.0005*	*0.0268*	*0.0005*	*0.0004*
40d/171d	0.3934	0.2121	*<0.0001*	*0.0003*	0.1009	0.8986	*0.0028*	*0.0001*	*0.0003*	*0.0009*	*0.0049*	*0.0003*	*0.0003*
70d/90d	0.8589	0.1145	0.8458	0.8348	0.8927	0.8053	0.7238	0.7404	0.0828	0.7064	0.8428	0.4343	0.6528
70d/107d	0.5665	*0.0273*	*0.0032*	*0.0028*	0.1519	0.2153	*0.0007*	*0.0013*	*0.0074*	*0.0004*	0.7198	*0.0170*	*0.0086*
70d/121d	0.6638	*0.0346*	*0.0005*	0.0887	*0.0210*	0.7178	*0.0038*	*0.0071*	*<0.0001*	*0.0003*	0.9756	*0.0280*	*0.0165*
70d/171d	0.7277	0.0635	*0.0014*	0.0578	*0.0208*	0.1842	*0.0014*	*0.0064*	*0.0002*	*0.0006*	0.4182	*0.0133*	*0.0117*
90d/107d	0.6916	0.4480	*0.0048*	*0.0018*	0.1898	0.3129	*0.0014*	*0.0007*	0.2290	*0.0009*	0.8721	*0.0035*	*0.0212*
90d/121d	0.7968	0.5221	*0.0008*	0.0605	*0.0275*	0.9080	*0.0077*	*0.0036*	*0.0027*	*0.0007*	0.8668	*0.0058*	*0.0400*
90d/171d	0.8644	0.7448	*0.0021*	*0.0389*	*0.0271*	0.2711	*0.0028*	*0.0033*	*0.0068*	*0.0012*	0.3177	*0.0027*	*0.0286*
107d/121d	0.8890	0.9030	0.3682	0.0959	0.2972	0.3686	0.4130	0.3954	*0.0318*	0.8835	0.7427	0.8014	0.7458
107d/171d	0.8209	0.6606	0.6854	0.1438	0.2941	0.9226	0.7428	0.4228	0.0763	0.8678	0.2500	0.9021	0.8791
121d/171d	0.9307	0.7504	0.6138	0.8163	0.9944	0.3213	0.6187	0.9601	0.6449	0.7547	0.4014	0.7085	0.8630

[a]ANKRD2: ankyrin repeat domain 2, MYH2: myosin, heavy chain 2, ACTA1: actin, alpha 1, DCTN1: dynactin 1, MYBPC1: myosin binding protein C, NEB: nebulin, TPM2: tropomyosin, 2, OBSCN: obscurin, ANK1: ankyrin 1, MYH7: myosin, heavy chain 7, MYL2: myosin, light chain 2, TNNT3: troponin T, type 3, TP53: tumour protein, p53.

[b]21d, 40d, 70d and 90d are days of pregnancy and 107d, 121d and 171d are days after birth. Values in italic format were statistically significant with a p value lower than 0.05.

muscle system process and myofibril subnets. The TPM2 gene in human is expressed in both pre- and postnatal stage (Wang *et al.* 2003), but our results showed higher expression during foetal growth going down at postnatal stage, demonstrating a differential of expression between them. Other genes known to be related with myogenic stage as ANKRD2 (Bean *et al.* 2008), MYBPC1 (Gautel *et al.*, 1998 and Kurasawa *et al.* 1999), NEB (Bang *et al.* 2006) and MYL2 (Zhang *et al.* 2009) also showed a prenatal high expression in this study and are connected by muscle system process, structural constituent of muscle and myofibril subnets at the major network.

The computational approaches allowed the identification of genes being differentially expressed by their functional annotation. It was possible to find genes with obvious roles in muscle physiology such as those that encode proteins of cytoskeleton allowing inserting these genes into relevant functional networks. These can be helpful to better understand the molecular mechanisms responsible for the process of muscularity. In this work, the functional analysis results suggest an interaction between 21 DEGs mapped in the same genetic network related to skeletal muscle development. The present study also provides rich new information resource about genes that are cited to be involved in myogenesis for the first time, increasing our understanding about the molecular mechanisms underlying pig skeletal muscle development (TP53 and DCTN1). Here, they showed a high expression level during the first muscle development wave, which is known to be the primary generation stage of muscle formation and is connected with Wnt signalling pathway and motor activity subnets, respectively. This is helpful to better understand the molecular basis involved with skeletal muscle hyperplasic and hypertrophic mechanisms. However, this is a complex trait regulated by several transcription factors, many of them still to be identified.

Our comparative analysis of the prenatal and postnatal skeletal muscle shows genes, which play important role for one stage more than to another as discussed above. This finding can contribute to better explain gene function dynamics during muscle development and growth. More generally, our data are likely to be helpful in uncovering the pathways that

mediate prenatal and postnatal muscle development in vertebrates. A number of DEGs were identified across stages associated with meat production traits, which may be commercially valuable and even targets for candidate gene prospection and marker-assisted selection.

Acknowledgements

We would like to thank CNPq, CAPES, INCT-CA and FAPEMIG for financial support.

References

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Audic S., Claverie J.M. (1997) The significance of digital gene expression profiles. *Genome Res.*, **10**, 986–995.

Bang M.L., Li X., Littlefield R., Bremner S., Thor A., Knowlton K.U., Lieber R.L., Chen J. (2006) Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle. *J Cell Biol.*, **173**, 905–916.

Bean C., Facchinello N., Faulkner G., Lanfranchi G. (2008) The effects of Ankrd2 alteration indicate its involvement in cell cycle regulation during muscle differentiation. *Biochim. Biophys. Acta – Mol. Cell Res.*, **1783**, 1023–1035.

Eisen M.B., Spellman P.T., Brown P.O., Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.

Foreign Agricultural Service/USDA Office of Global Analysis, (2011) (available at: http://www.fas.usda.gov/livestock_arc.asp; last accessed 5 October 2011).

Gautel M., Fürst D.O., Cocco A., Schiaffino S. (1998) Isoform transitions of the myosin binding protein C family in developing human and mouse muscles: lack of isoform transcomplementation in cardiac muscle. *Circ. Res.*, **82**, 124–129.

Gorodkin J., Cirera S., Hedegaard J., Gilchrist M.J., Panitz F., Jørgensen C., Scheibye-Knudsen K., Arvin T., Lumholdt S., Sawera M., Green T., Nielsen B.J., Havgaard J.H., Rosenkilde C., Wang J., Li H., Li R., Liu B., Hu S., Dong W., Li W., Yu J., Wang J., Stærfeldt H.-H., Wernersson R., Madsen L.B., Thomsen B., Hornshøj H., Bujie Z., Wang X., Wang X., Bolund L., Brunak S., Yang H., Bendixen C., Fredholm M. (2007) Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. *Genome Biol.*, **8**, R45.

Greller L.D., Tobin F.L. (1999) Detecting selective expression of genes and proteins. *Genome Res.*, **3**, 282–296.

Kurasawa M., Sato N., Matsuda A., Koshida S., Totsuka T., Obinata T. (1999) Differential expression of C-protein isoforms in developing and degenerating mouse striated muscles. *Muscle Nerve*, **22**, 196–207.

Lie D.C., Colamarino S.A., Song H.J., Désiré L., Mira H., Consiglio A., Lein E.S., Jessberger S., Lansford H., Dearie A.R., Gage F.H. (2005) Wnt signalling regulates adult hippocampal neurogenesis. *Nature*, **437**, 1370–1375.

Murray D., Doran P., MacMathuna P., Moss A.C. (2007) *In silico* gene expression analysis an overview. *Mol. Cancer*, **6**, 50.

Nascimento C.S., Peixoto J.O., Verardo L.L., Campos C.F., Weller M.M.C., Faria V.R., Botelho M.E., Martins M.F., Machado M.A., Silva F.F., Lopes P.S., Guimarães S.E.F. (2012) Transcript profiling of expressed sequence tags from semimembranosus muscle from Commercial and native breeds of pig. *Genet. Mol. Res.*, **11**, 3315–3328.

Pfaffl M.W. (2001) A new mathematical model for relative quantification in real time RT-PCR. *Nucleic Acids Res.*, **29**, e45.

Radoniæ A., Thulke S., Mackay I.M., Landt O., Siegert W., Nitsche A. (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochem. Biophys. Res. Commun.*, **313**, 856–862.

Rehfeldt C., Fiedler I., Dietl G., Ender K. (2000) Myogenesis and postnatal skeletal muscle cell growth as influenced by selection. *Livest. Prod. Sci.*, **66**, 177–188.

Rehfeldt C., Fiedler I., Stickland N.C. (2004) Number and size of muscle fibers in relation to meat production. In: M.F.W. te Pas, M.E. Haagsman, H.P. Everts (eds), Muscle Development of Livestock Animals: Physiology, Genetics, and Meat Quality. CAB International, Wallingford, pp. 1–38.

Romualdi C., Bortoluzzi S., Danieli G.A. (2001) Detecting differentially expressed genes in multiple tag sampling experiments. Comparative evaluation of statistical tests. *Hum. Mol. Genet.*, **10**, 2133–2141.

Romualdi C., Bortoluzzi S., D'Alessi F., Danieli G.A. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol. Genomics*, **12**, 159–162.

Różycki M., Piórkowska K., Oczkowicz M. (2011) Expression of contractile protein-encoding genes TPM2 and TNNT3 during ontogenesis in pigs raised in Poland. *Anim. Sci. Pap. Rep.*, **29**, 65–74.

Ryu Y.C., Lee M.-H., Lee S.-K., Kim B.C. (2006) Effects of muscle mass and fiber type composition of *Longissimus dorsi* muscle on postmortem metabolic rate and meat quality. *J. Muscle Foods*, **17**, 343–352.

Serão N., Veroneze R., Ribeiro A., Verardo L.L., Braccini N.J., Gasparino E., Campos C.F., Lopes P.S., Guimarães S.E.F. (2010) Candidate gene expression and intramuscular fat content in pigs. *J. Anim. Breed. Genet.*, **128**, 28–34.

Sollero B.P., Guimarães S.E.F., Rilington V.D., Tempelman R.J., Raney N.E., Steibel J.P., Guimarães J.D., Lopes P.S., Lopes M.S., Ernst C.W. (2011) Transcriptional profiling during foetal skeletal muscle development of Piau and Yorkshire–Landrace cross-bred pigs. *Anim. Genet.*, **42**, 600–612.

Stekel D.J., Git Y., Falciani F. (2000) The comparison of gene expression from multiple cDNA libraries. *Genome Res.*, **12**, 2055–2061.

Steibel J.P., Poletto R., Coussens P.M., Rosa G.J.M. (2009) A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics*, **94**, 146–152.

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A., Speleman F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, RESEARCH0034.

Wang H., Hubbell E., Hu J.-s., Mei G., Cline M., Lu G., Clark T., Siani-Rose M.A., Ares M., Kulp D.C., Haussler D. (2003) Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics*, **19**, 315–322.

Wigmore P.M.C., Stickland N.C. (1983) Muscle development in large and small pig fetuses. *J. Anat.*, **137**, 235–245.

Zhang S.-Z., Xu Y., Xie H.Q., Li X.-Q., Wei W.-Q., Yang Z.-M. (2009) The possible role of Myosin light chain in myoblast proliferation. *Biol. Res.*, **42**, 121–132.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Nucleotide sequence of the primers, accession number and amplicon length of nucleotide sequences used in Real-Time PCR reactions identified as differentially expressed.

Table S2. Deferentially Expressed Genes primer concentration and cDNA. Their concentrations were determinate through amplification efficiency as described by Pfaffl 2001.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.