LUCAS LIMA VERARDO

DIFFERENTIALLY EXPRESSED GENES AND miRNA IDENTIFICATION IN PIG SKELETAL MUSCLE

Dissertation presented to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRAZIL
2011

LUCAS LIMA VERARDO

DIFFERENTIALLY EXPRESSED GENES AND miRNA IDENTIFICATION IN PIG SKELETAL MUSCLE

Dissertation presented to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of Magister Scientiae.

APROVADA: 25 de julho de 2011.

Marta Fonseca Martins Guimarães

Paulo Sávio Lopes
(Co-orientador)

Simone Eliza Facione Guimarães
(Orientadora)

AKNOWLEDGEMENTS

Firstly, I would like to thank God for the care and comfort in every moment and for giving me strength to finish this study.

Sincere thanks also to:

The most important people in my life, my parents Márcio and Beth, who always supported me and believed in me;

My brother and sister (Paulo and Liana) and my girlfriend Juliana for giving me all support and friendship;

Friends of República Inçana (Thiago, Marcos, Pedrão, Pedrinho, Alan, Zé, Renê, Matheus, João, Gaucho, Renan, Vinicius and Guilherme) and friends of Judô team (Rangel, Mococa, Maguinho, Thiagão, Allan, Samuel and etc.) for providing me some of the best years of my life here in Viçosa;

My friends from LABTEC (Animal Biotechnology Laboratory), Kleibe, Marcos, Nicola, Erika, Mayara, Priscila, Renata, Ana Paula, Kati, Bruna, Débora, Carol, Margareth, Iuri, André, Daniel, Darlene and Beatriz for continuous knowledge, learning exchange and pleasant workplace;

Federal University of Viçosa (UFV), in special to the Animal Science Department (DZO) and Genetics and Breeding Program, for the opportunity of carrying out the course;

Professor Simone Eliza Facioni Guimarães, my adviser, for her excellent supervision, competence, teachings, friendship and opportunities given to me;

Dr. Carlos, my co-adviser for him patience, friendship and teaching, a lot of teaching;

Professor Paulo Sávio Lopes, for his valuable teachings, patience and friendship;

Professor Fabyano Fonseca for all patience, his help during analyses and SAS lessons;

Dr. Marta Fonseca Martins, for participate of the review of my Dissertation and for her presence;

Foundation for the Coordination of High Education and Graduate Training (CAPES) for the financial support;

TABLE OF CONTENTS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

RESUMO

VERARDO, Lucas Lima, M.Sc., Universidade Federal de Viçosa, Julho de 2011. **Identificação de genes diferencialmente expressos e miRNAs em músculo esquelético de suínos.** Orientadora: Simone Eliza Facioni Guimarães. Co-orientadores: Paulo Sávio Lopes e Carlos Souza do Nascimento.

O suíno (*Sus scrofa*) é considerado um animal de grande importância para produção de carne, sendo seu potencial de crescimento muscular objeto de grande interesse e geralmente associado com características determinadas na fase pré-natal durante a miogênese. Para o estudo de genes responsáveis por estas características, as etiquetas de sequências expressas (*Expressed Sequence Tags* - EST) fornecem informações diretas sobre o transcriptoma e indiretas sobre a relação entre o genoma e diferentes fenótipos, proporcionando o conhecimento sobre genes diferencialmente expressos (GDE) bem como sequências genômicas transcritas para o controle da expressão gênica como, por exemplo, alguns RNAs não codificantes. Características de tecidos musculares em suínos podem ser influenciadas diretamente por genes, e estes sendo regulados como, por exemplo, através de miRNAs, em diferentes fases de desenvolvimento. O presente trabalho teve como objetivo a identificação e a anotação *in sílico* de GDE e sequências não codificantes, com enfoque aos miRNAs, de bibliotecas de cDNA construídas a partir do músculo esquelético semi-membranoso de três diferentes raças de suínos (Duroc, Large White e naturalizada brasileira Piau) bem como a análise dos níveis de expressão dos genes identificados e miRNAs em sete fases de desenvolvimento do *Longissimus Dorsi* (21, 40, 70 e 90 dias pré-natal e 107, 121 e 171 dias pós-natal) de animais de linha comercial. Foram identificados 34 GDE sendo 21 pertencentes a uma rede gênica musculo-específica. Destes, 13 genes tiveram seus perfis de expressão analisados com o uso do qRT-PCR durante os sete períodos citados, formando quatro grupos de expressão semelhantes, um com maior expressão na fase pós-natal e três na fase pré-natal. Nas análises das sequências não codificantes um resultado importante foi a identificação de dois novos miRNAs em suínos, os quais tiveram suas sequências maduras similares aos miRNAs hsa-miR-1207-5p e hsa-miR-665 foram classificadas como verdadeiras pelo programa MiPred e formaram estruturas secundárias. Destes, encontrou-se 289 e 214 genes regulados por eles respectivamente, dos quais quatro são músculo-específicos. Os novos miRNAs tiveram seus perfis de expressão analisados com o uso do PCR em tempo real durante os sete períodos citados juntamente com outros três já identificados em suínos. Seus níveis de expressão mostraram diferenças entre os estágios pré- e pós-natal. Estes estudos podem fornecer valiosas informações possibilitando um maior entendimento dos mecanismos moleculares envolvidos no desenvolvimento muscular. As análises de GDE em fases pré e pós-natal sugerem a presença de genes atuando especificamente em determinados estágios de desenvolvimento do músculo, contribuindo para melhor explicar suas funções. A identificação de dois novos miRNAs, somados a outros já identificados e postados nos bancos de dados em suínos, podem contribuir para um maior entendimento dos modos de regulação gênica, sendo de importância para os estudos de genética e melhoramento animal, permitindo o entendimento da fisiologia da deposição de músculo para produção de carne em suínos.

ABSTRACT

VERARDO, Lucas Lima, M.Sc., Universidade Federal de Viçosa, July, 2011. **Differentially expressed genes and miRNA identification in pig skeletal muscle.** Adviser: Simone Eliza Facioni Guimarães. Co-Advisers: Paulo Sávio Lopes and Carlos Souza do Nascimento.

The pig (*Sus scrofa*) is considered an important animal for meat production. This interest revolves around the potential for muscle growth, which usually is associated with certain characteristics during prenatal myogenesis. To study the genes responsible for these characteristics, expressed sequence tags (EST) provide direct information about the transcriptome and indirectly on the relationship between the genome and different phenotypes, supplying knowledge about differentially expressed genes (DEG) as well as other transcribed genomic sequences for the control of gene expression, e.g., some non-coding RNAs. Characteristics of muscle tissue in pigs may have been directly influenced by genes, and those being regulated, for example, by miRNAs, in different stages of development. This study aimed to identify by *in silico* annotation, DEG and non-coding sequences, focusing on miRNAs, using cDNA libraries constructed from semi-membranous skeletal muscle of three different pig breeds (Duroc, Large White and naturalized Brazilian Piau) as well as analysis of gene expression profiles of identified genes and miRNAs during seven stages of development (21, 40, 70 and 90 days prenatal and 107, 121 and 171 days postnatal) from commercial line animals *Longissimus Dorsi* muscle. Twenty-one identified genes out of 34 DEGs belongs to the muscle-specific path. From these, 13 genes had their expression profiles analyzed by qRT-PCR during the seven periods, forming four clusters of similar expression, with one having greater expression in the postnatal period and three in the prenatal. In the analysis of non-coding sequences, an important result was the identification of two new miRNAs in pigs, which had their sequences similar to mature miRNAs hsa-miR-1207-5p and hsa-miR-665 which had their precursor sequences forming secondary structures and classified as real precursor sequence by MiPred program. From these, we found 289 genes and 214 respectively regulated by them, of which four are muscle-specific. The new miRNAs and other three which have been identified in previous studies in pigs had their expression levels analyzed by quantitative real time PCR during the mentioned seven periods. Their levels of expression differed between pre- and postnatal stages. These studies may provide valuable information allowing a better understanding of the molecular mechanisms involved in muscle development. Analyses of DEG in the pre-and postnatal periods suggest the presence of genes acting specifically on certain stages of muscle development, contributing to better explain their functions. The identification of two new miRNAs, together with other previously identified and posted on the databases in pigs, may contribute to a better understanding of gene regulation and is important for studies of genetics and animal breeding, allowing the understanding of the muscle deposition physiology to meat production in pigs.

Chapter 1

GENERAL INTRODUCTION

Muscle development

The interest in muscle growth potential is generally associated with characteristics determined prenatally during myogenesis (Rehfeldt et al., 2000). Recent studies in fetal 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 looking for a better meat productivity and quality still have a particular place in the area of animal genetics research.

The lean muscle growth potential of an animal largely depends on the number of muscle fibers prenatally formed, once the postnatal increase in muscle fiber size is limited by genetic and physiological reasons (Rehfeldt et al., 2004), so the skeletal muscle development is an important process in meat animals, this directly affects meat production. The animal muscle mass is formed in two stages during gestation, including primary and secondary fiber formation, being the muscle fiber number fixed before birth. A primary generation occurs from 35 until 55 days of gestation (dg), followed by a second generation which happens between 55 and 90-95 dg. These secondary fibers form around the primary myotubes, using them as a scaffold. The total number of muscle fibers is generally considered to be definitively established by 90-95 dg in the pig and birth occurring at 113 dg (Lefaucheur et al., 1995).

During myogenesis stage there are some muscle-specific genes which may play roles on that. Examples are: ANKRD2, contributing to the coordination of proliferation and apoptosis during myogenic differentiation in mouse (Bean et al., 2008); MYBPC1 which is known to be expressed in skeletal muscle prior to birth in human and mouse (Gautel et al., 1998 and Kurasawa et al., 1999); NEB, proposed to play critical roles in myofibrillogenesis (Bang et al., 2006) and MYL2 which has been speculated to be involved in myogenesis or muscle regeneration through its family function in myogenic cells (Zhang et al., 2009).

One of the unique characteristics of the skeletal muscle is the diversity on its morphological and biochemical properties (Ryu et al., 2006), what may be explained by the type of protein present according to the expressed genes. Investigation of expressed genes during skeletal muscle development is elementary to understand molecular mechanism of muscle growth. The pattern of expression can be evaluated by analyses of expressed sequence tags (EST) present in muscle tissues. 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 wide range of applications (Murray et al. 2007), for example to identify genes of importance in meat production or to analyze the effect of a specific gene in the muscle development. Nevertheless, the simple information of these genes is not enough to understand their role in muscularity. There are many factors acting on the genes expression and a good example would be non-coding RNA sequences.

Non-coding RNA

In a first moment the name non-coding RNA (ncRNA) give us an idea that such RNAs neither contain information nor have function. Although it has been generally assumed that most genetic information is transacted by proteins, recent evidence suggests that the majority of mammal's genomes and other complex organisms are in fact transcribed into ncRNAs, many of which are alternatively spliced and/or processed into smaller products (Jacquier, 2009). Some of these ncRNAs include micro RNAs which are a class of naturally occurring RNA, first recognized in plants and worms and have been extensively characterized since then in other species (Watson et al., 2004).

Typically miRNA have ~22 nucleotides (nt) in length that suppress translation of their targets genes by binding to the 3' untranslated region (Ambros, 2004 and Bartel, 2004). Recent studies have shown that miRNAs play important gene-regulatory roles in numerous eukaryotic lineages and are often highly conserved across animal species. They are involved in many diverse biological processes and may potentially regulate the functions of thousands of genes. The regulation mechanisms can be since the repression of translation to the cleavage of targeted mRNAs. This class of RNA may directly target transcription factors which affect animal development, and also specific genes which control metabolism (Carthew, 2006). Identification of comprehensive miRNAs sets and other small regulatory RNAs in different organisms is a critical step to facilitate our understanding of genome organization, genome biology and evolution (Carrington & Ambros, 2003).

Characteristics of muscle tissue in pigs may have been directly influenced by genes, and those being regulated, for example, by miRNAs, in different stages of development. The goal of this work was to identify genes and miRNAs which have been expressed in pig breeds differing in muscularity and analyze is their expression in different muscle development periods in commercial animals.

REFERENCES

Ambros, V. **The functions of animal microRNAs**. *Nature* 2004, 431: 350–355.

Bartel, DP. **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**, 281−297.

Carrington, JC; Ambros, V. **Role of microRNAs in plant and animal development**. *Science* 2003, **301**(5631): 336–338

Carthew, RW. **Gene regulation by microRNAs.** *Current Opinion in Genetics & Development* 2006, **16,** 203–208.

Gorodkin J., Cirera S., Hedegaard J. et al. **Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags.** *Genome Biology* 2007, **8:**R45

Jacquier, A. **The complex eukaryotic transcriptome: unex- pected pervasive transcription and novel small RNAs.** *Nature Reviews Genetics* 2009, **10**, (12):833–844.

Lefaucheur, L., Edom, F., Ecolan, P. & Butler-Browne, G.S. **Pattern of muscle fiber type formation in the pig**. *Developmental Dynamics* 1995, **203**, 27-41.

Murray D, Doran P, MacMathuna P, Moss AC: ***In silico* gene expression analysis – an overview.** *Molecular Cancer* 2007, **6**:50.

Rehfeldt, C, I Fiedler, G Dietl, and K Ender: **Myogenesis and postnatal skeletal muscle cell growth as influenced by selection.** *Livestock Production Science* 2000, **66**:177–188.

Rehfeldt, C, I Fieldler and NC Stickland: **Number and Size of Muscle Fibres in Relation to Meat Production.** In: MFW te Pas, ME Everts and HP Haagsman, *Muscle development of livestock animals: physiology, genetics, and meat Quality* 2004, p. 15 – 52. CABI Publishing: Oxon.

Ryu, YC, M-H Lee, S-K Lee and BC Kim: **Effects of muscle mass and fiber type composition of *Longissimus dorsi* muscle on postmortem metabolic rate and meat quality.** *Journal of Muscle Foods* 2006 17: 343-352.

Sollero BP, Guimarães SEF, Rilington V D, Tempelman R J, Raney N E, Steibel J P, Guimarães J D, Lopes P S, Lopes M S and Ernst C W. **Transcriptional profiling during foetal skeletal muscle development of Piau and Yorkshire–Landrace cross-bred pigs**. *Animal Genetics* 2011, doi:10.1111/j.1365-2052.2011.02186.x

Swatland, HJ (1994) In**: Structure and Development of Meat Animal and Poultry**. Technomic Publishing Co., Inc., Lancaster, PA, P. 303.

Watson JD, Baker TA, Bell SP, Gann A et al. (2004). **Molecular Biology of the Gene.** 5th edition. Pearson Education, Inc, publishing as Benjamin Cummings, 1301 Sansome Street, San Francisco.

Chapter 2

DIFFERENTIALLY EXPRESSED GENES IDENTIFIED FROM ADULT PIG cDNA LIBRARIES AND THEIR EXPRESSION LEVELS.

Abstract. Pig (*Sus scrofa*) is an important animal, for meat production as well as model organism for biomedical research. As animal production, an interest in the muscle growth potential is generally associated with characteristics determined prenatally during myogenesis. Expressed sequence tags (EST) and full-length cDNAs provide direct information on the transcriptome and indirect information on the relation between the genome and different phenotypes, giving information about differentially expressed genes (DEG). In this work the *in sílico* identification and annotation of DEG from EST libraries of three different pig breeds (Duroc, Large White and Brazilian naturalized Piau) were performed, followed by qRT-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 the analysis of genes expression levels across these periods. Thirty four genes differentially expressed were identified, being 21 grouped in a network related to muscle traits which is composed by subnets like myofibril, muscle system process, structural constituent of muscle, motor activity and Wnt signaling pathway after annotation. Thirteen had their expression profile measured by qRT-PCR during all seven periods and could be grouped in four cluster according to their expression level having genes being more expressed during prenatal (ANKRD2, MYBPC1, NEB, DCTN1, MYL2, TP53, TPM2, OBSCN and ANK1) and others during postnatal stage (MYH2, ACTA1, MYH7 and TNNT3). These findings can contribute to better explaining genes function mechanism. More generally, our data are likely to be helpful in uncovering the pathways that mediate prenatal and postnatal skeletal muscle development in vertebrates.

Keywords: expressed sequence tag, *in sílico*, prenatal, qRT-PCR.

Introduction

The 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 fetal 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 area 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 fibers prenatally formed, being the postnatal increase in muscle fiber 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 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 analyze 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 quantitative real time PCR.

Here we made use of our results from an *in sílico* identification of differentially expressed genes in EST libraries of three different pig breeds (Duroc, Large White and Brazilian naturalized Piau) to annotate their metabolic path and select differentially expressed genes related to muscularity to analyze through qRT-PCR their expression profile in pre- and postnatal stages from commercial breed animals.

Material and Methods

All methods involving animal handling was done 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

As a multidimensional analysis of data from Expressed Sequence Tags (ESTs), the *in sílico* differential expression methodology allow us to identify differentially expressed genes (DEGs) in cDNA libraries (Audic and Claverie, 1997). All ESTs used in this study were generated from three different pig breeds semimembranosus skeletal muscle tissue cDNA libraries (Duroc, Large White and a naturalized Brazilian breed, Piau) constructed by Nascimento et al. (2011).

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 with the parameters set greater than 80% similarity over at least a 40-bp fragment. ESTs that did not form contigs were designated as singletons. The number of sequences on each contig was determined and used to construct a table for further analysis. Functional annotation of ESTs for the three sets of data was performed using the program BlastX (Altschul et al., 1990) using an e-value $\leq 10^{-05}$, against the data base of transcripts from SwissProt (http://www.expasy.ch/sprot/). 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, 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 differentially expressed genes in multiple tag sampling experiments. For pair-wise comparisons, we applied the AC statistic (Audic and Claverie, 1997), Fisher's 2x2 exact test and 2x2 Chi^2 test, and in multiple comparisons the R statistic (Stekel et al., 2000), GT statistic (Greller and Tobin, 1999), and general Chi^2 test were used. For each test the associated significance thresholds (<0.05) were applied using the Bonferroni correction and the differentially expressed genes were identified.

Gene Network Analysis

Aiming to examine the process of shared pathways, the Ensembl gene identifiers can be 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 differentially expressed genes. The application Cytoscape (www.cytoscape.org/) was used to visualize and edit the identified pathways. Only networks that include direct relationships between genes and muscularity have been maintained. Based on the functional genes annotated as differentially expressed, we obtained a list (**Table 1**) of 13 candidate genes for experimental validation by qRT-PCR.

Quantitative real-time PCR of differentially expressed genes

Tissue samples

A total of 21 *Longissimus dorsi* (LD) muscle samples divided in both, pre- and postnatal stages were collected. At the UFV Pig Breeding Farm, pregnant Commercial gilts at 21, 40, 70 and 90 days of gestation were aborted using the following protocol: intramuscular injections of 1 mL Prelobam® (PGAF-α)-plus 1 mL Estrogen, followed 12 h later by 2 mL Orastina® (Ocitocine). LD muscle samples were isolated from 12 fetuses, 21d (n=3), 40d (n=3), 70d (n=3) and 90d (n=3); and placed in sterile tubes containing RNA*later*® (Qiagen). Samples were stored at 4ºC overnight and at -70ºC prior to RNA isolation. In the same way, samples of LD from Commercial castrated males in three different ages, 107 days (n=3), 121 days (n=3) and 171days (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 NanoVue^{TM} Plus (GE Healtchare, Germany) and quality at the Agilent 2100 Bioanalyzer © (Agilent Technologies, Palo Alto, Calif.) obtained an 7.6 average RIM quality value. RNA samples were kept at -70^{o} C until cDNA synthesis reactions (reverse transcriptase). More details about the animals and RNA extraction were described by Serão et al. (2010) and Sollero et al. (2010).The first strand of cDNA synthesis was performed using ProtoScript® M-MuLV First-Strand cDNA Synthesis Kit (New England Biolabs Inc., Beverly, MA) and its concentration was estimated in a spectrophotometer NanoVue TM Plus (GE Healtchare, Germany). Subsequently, the cDNA samples were kept at -20 °C until the completion of the PCR reactions in real time.

Quantitative real-time PCR analysis

The qRT-PCR reactions were performed in thermal cycler ABI Prism 7300 Sequence Detection Systems (Applied Biosystems, Foster City, CA, USA) using GoTaq® qPCR Master Mix (Promega Corporation, Madison, USA). Amplification conditions for all systems were: 95 °C for 2 minutes, 40 cycles of denaturation at 95 °C for 15 seconds, and extension at 60 °C for 60 seconds. After 40 cycles of amplification, an additional step with a gradual increase in temperature of 60 to 95 °C was used to obtain the dissociation curve. The best primers and cDNA amplification conditions were used for qRT-PCR analyzes. The amplification of target genes was performed at different wells and in duplicates (Livak and Schmittigen, 2001). The concentrations of primers and cDNA used in amplification reactions are shown in **Table S1**. The used experimental design was completely randomized, with three replicates 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 qRT-PCR were analyzed 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} ~N(0, σ^2_D); and E_{ijk} is a residual term., e_{ijk} ~ N(0, σ^2_e).

All statistical procedures were performed using SAS 9.0 for Windows (Statistical Analysis System Institute, Inc., Carry, NC, USA). The routine QPCR_MIXED: SAS® [https://www.msu.edu/~steibelj/JP files/QPCR.html] developed to generate commands in SAS PROC MIXED used to analyze qRT-PCR data was used 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 (SAS software) using t-student test.

Gene target sequence was recovered from nucleotide sequences obtained from the GenBank database (www.ncbi.nlm.nih.gov). These sequences were used to construct primers by the PrimerQuest program available at (www.idtdna.com/SciTools/Applications/primerQuest) provided by Integrated DNA Technologies, Inc (Coralville, IA). Nucleotide sequence of the primers and accession numbers of nucleotide sequences are summarized in **Table 1**.

Table 1. Nucleotide sequence of the primers and accession number of nucleotide sequences used in qRT-PCR reactions identified as differentially expressed across the EST libraries analyzed.

Genes	Nucleotides Sequences (5' - 3')	Access ID*	Amplicon Length (nt)
ANKRD2	F: GATCCAGAACCTCATCGAGC R: ACCGCAGCTTTCAGGAAG	NM_020349	138
MYH2	F: AGTAGGAAAGTCGCAGAACAG R: CCATTTCTCCTTGCATTTGGG	AB_025260.1	133
ACTA1	F: GGTCGGTATGGGTCAGAAAG R: TTCTCCATGTCATCCCAGTTG	NM_001100.3	120
DCTN1	F: CGGAACCTGAATCTGGAAGAG R: TGCAGCTCATCGTTCATCTC	NM_004082	89
MYBPC1	F: CTATTCTCAGCCCATTCTCGTG R: TCTGGTCTTGGTTTTCCCTG	NM_002465.2	141
NEB	F: AGGAAGCAATAGGACAAGGAAC R: CAATCTCTGGAGTCACAGTGG	NM_004543.3	141
TPM2	F: GACAGAGGATGAGGTGGAAAAG R: GTTCAGGGAGGCCACATC	NM_003289	100
OBSCN	F: CACCTTTGTGCAAGAGCGTGAGAA R: AAGAGATATGTCCCAGGAGTGCCA	XM_003123629.1	101
ANK1	F: TCTTGAGCCCTGTGATTGTG R: CAGGTAGCTTTCTCCATAGCG	NM_021009.4	134
MYH7	F: CTGCTTCGGGAACAGTATGAG R: TGGCATCTGTCTCGTATTTGG	NM_001100.3	115
MYL2	F: ACCATGTCACCTAAGAAAGCC R: ATGAAGCCATCCCTGTTCTG	NM_000037.3	134
TNNT3	F: ATTGACAAGTTCGAGTACGGG R: CACGTTACTTCCACCTTCCG	NM_020349	148
TP53	F: CGTCCTTTGTCCCTTCTCAG R: AACAGCTTATTGAGGGCAGG	NM_004082	121
TTN**	F: TGCTGGTTTCTATGTGGTCTG R: TCGCTAACTTTGACACCTCTG	XM_001925803.1	111

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- tumor protein, p53 and TTN- titin * Access ID: GenBank database identification (www.ncbi.nlm.nih.gov). ** Internal control gene.

Clustering relative quantification of gene expression

Hierarchical clustering was performed on qRT-PCR data (Eisen et al. 1998). ΔCt values (target Ct – endogenous Ct) were used in the analysis (Dittmer et al. 2005). The set of muscle-specific genes expression values across seven periods was used for unsupervised hierarchical 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. Three thousand six hundred and seventy 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. After BlastX of all 905 contig against SwissProt data base and further IDEG6 analysis, a total of 54 differentially expressed genes (DEGs) were identified, 34 represented coding genes for known proteins as following on supplementary data (**Table S2**).

Gene Networks Analyses

To understand the functions of 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 others related ones. Among 34 genes analysed, 16 genes (MYH2, DCTN1, PDLIM5, TPM2, MYBPC1, TNNT3, ACTA1, MYL2, MYH1, MYH7, SMARCD3, ANKRD2, MYLPF, ANK1, TRIM54 and MYBPH) and other five genes added due to their presence at the same metabolic path (MYH4, TP53, TNNC1, NEB and OBSCN) could be grouped into a network of functional relevance for muscularity. 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 signaling pathway. The latter subnet is a network of proteins best known for their roles in embryogenesis and cancer, but also involved in normal physiological processes in adult animals (Lie et al., 2005), where seven proteins on that net were found (SMARCD3, ACTA1, TP53, MYH1, MYH2, MYH4 and MYH7). In addition, this major network is composed of five secondary networks connected to each other. These are: myofibril, muscle system process, structural constituent of muscle, motor activity and Wnt signaling pathway. **Table 2** describes the function of the DEGs which are present on the network using the assignment of GO terms.

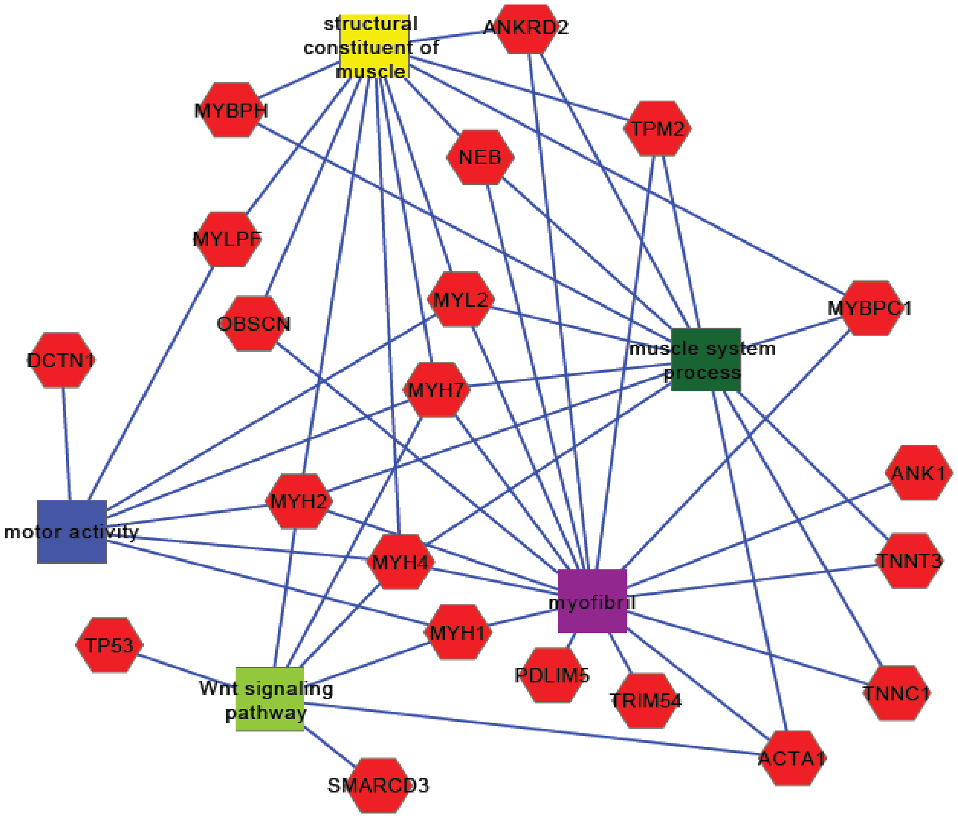


Figure 1. Functional gene networks and their interactions. It describes the relationships between 16 differentially expressed genes and other five included (in red), there are five important muscle development related subnets; motor activity (in blue), structural constituent of muscle (in yellow), muscle system process (in dark green), myofibril (in purple) and Wnt signaling pathway (in light green).

Table 2. Pathway, biological process, molecular function and cellular component from genes represented in the network. There are 21 genes related with muscle tissue development.

Genes	Pathway	GO: Biological Process	GO: Molecular Function	GO: Cellular Component
MYL2*	Striated Muscle Contraction	muscle system process	structural constituent of muscle/motor activity	myofibril
MYBPC1*	Striated Muscle Contraction	muscle system process	structural constituent of muscle	myofibril
ACTA1*	Wnt signaling pathway	muscle system process	structural molecule activity	myofibril
MYH2*	Wnt signaling pathway	muscle system process	structural constituent of muscle/motor activity	myofibril
MYH7*	Wnt signaling pathway	muscle system process	structural constituent of muscle/motor activity	myofibril
TNNT3*	Striated Muscle Contraction	muscle system process	cytoskeletal protein binding	myofibril
TPM2*	Muscle contraction	muscle system process	structural constituent of muscle	myofibril
TP53*	Wnt signaling 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*	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 signaling pathway	muscle system process	structural constituent of muscle/motor activity	myofibril
PDLIM5	-	regulation of synaptogenesis	cytoskeletal protein binding	myofibril
MYH1	Wnt signaling pathway	-	motor activity	myofibril
TNNC1	Striated Muscle Contraction	muscle system process	cytoskeletal protein binding	myofibril
SMARCD3	Wnt signaling 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

*Genes used on qRT-PCR analyses.

Quantitative real-time PCR of differentially expressed genes

For a better understanding of the differentially expressed genes biological role, we performed a qRT-PCR across seven periods (four pre-natal and three post-natal). The efficiencies of amplification by qRT-PCR represented around 100% in each cycle, and the relative abundance was calculated using an equation to correct differences in efficiency as described by Pfaffl et al., 2001. 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 (te Pas et al. 2005 and 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 pattern of expression for each gene is reported by hierarchical clustering describing the expression level gene by gene to better evaluate differences between periods of each specific gene during embryo growth and adult age. During prenatal stage nine genes showed a higher expression 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 (ΔCt) across seven periods (21, 40, 70 and 90 days prenatal and 107, 121 and 171 days of adult age 1 to 7 respectively), have grouped all genes into four separate clusters. One containing only post-natal 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 ($p < 0.05$ and < 0.01), these data are showed in **Table 3**.

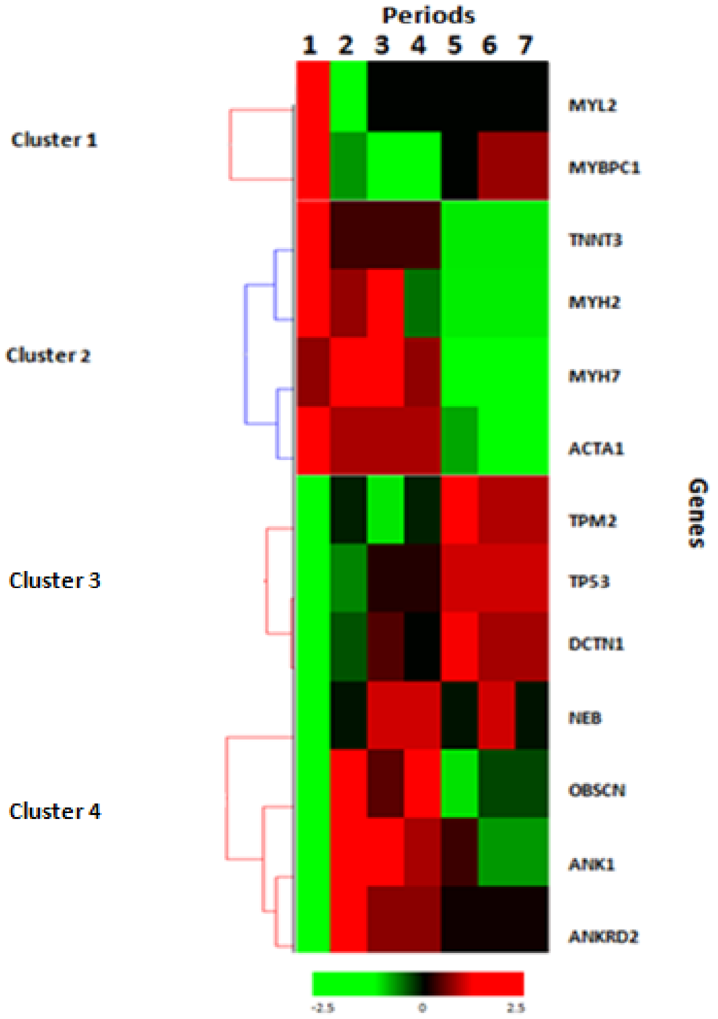


Figure 2. Unsupervised hierarchical clustering of muscle DEG expression levels across all seven periods (1-4 prenatal, 21, 40, 70 and 90 days of pregnancy respectively and 5-7 postnatal, 107, 121 and 171 days after birth). They were divided in four clusters by similar expression pattern. A median expression value equal to zero was designated in black; green increased expression and red reduced expression.

Table 3. p values for differential expression values for each gene in the comparison between 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	<.0001**	0.0027**	0.0022**	0.0682	<.0001**	<.0001**	0.2680	0.0010**	0.0002**	<.0001**
21d/70d	0.0099**	0.6030	0.0061**	<.0001**	0.0005**	0.0002**	0.1285	<.0001**	<.0001**	0.1841	0.1195	<.0001**	<.0001**
21d/90d	0.0151*	0.0438*	0.0041**	<.0001**	0.0007**	0.0003**	0.0682	<.0001**	<.0001**	0.3284	0.0840	<.0001**	<.0001**
21d/107d	0.0366*	0.0096**	<.0001**	<.0001**	0.0103*	0.0023**	<.0001**	0.1355	<.0001**	0.0065**	0.0625	<.0001**	<.0001**
21d/121d	0.0270*	0.0123*	<.0001**	<.0001**	0.0813	0.0004**	0.0002**	0.0275*	0.0013**	0.0048**	0.1133	<.0001**	<.0001**
21d/171d	0.0222*	0.0232*	<.0001**	<.0001**	0.0824	0.0028**	<.0001**	0.0303*	0.0005**	0.0091**	0.4237	<.0001**	<.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**	<.0001**	0.5160	0.9759	0.0014**	<.0001**	0.0137*	0.0007**	0.0505	0.0003**	0.0002**
40d/121d	0.3480	0.1250	<.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	<.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**	<.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- tumor protein, p53. **b**-21d, 40d, 70d and 90d are days of pregnancy and 107d, 121d and 171d are days after birth.* $p < 0.05$ and ** $p < 0.01$ statistical differences between periods.

Discussion

Comparison of three sequenced cDNA libraries from divergent genetic background identified sequences that were differentially expressed among them. The comparison of results derived from the six statistical tests showed that they produced different results, indicating divergence among them (**Table S3**). The Chi^2 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 analyzed 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^2 as the most appropriate combination to test deferential distribution in multiple tag sampling experiments with cDNA libraries.

For a better understanding of the differentially expressed genes biological role, we performed a qRT-PCR across seven periods (four pre-natal and three post-natal). During prenatal stage nine genes showed a higher expression in at least one period (ANKRD2, MYBPC1, NEB, DCTN1, MYL2, TP53, TPM2, OBSCN and ANK1) and four during postnatal (MYH2, ACTA1, MYH7 and TNNT3).

Prenatal over expressed

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 and Stickland, 1983). Here two genes (MYL2 and MYBPC1) had high expression in critical periods of pre-natal stage forming the Cluster 1 (**Figure 2**) 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.

Myosin light chains (MYLs) are the main components in myofibrils and are associated with myosin heavy chain heads. According to their conditions of dissociation, MYLs are divided into two classes (Barton et al., 1985; Hailstones et al., 1990). One is called the regulatory (or phosphorylatable) light chain *(i.e.* MYL2) and the other is the alkali light chain *(i.e.* MYL1, MYL3 and MYL4). Each class has several isoforms associated with different muscle types. It has been speculated that MYL may be involved in myogenesis or muscle regeneration through its function in myogenic cells (Zhang et al. 2009). The MYL2 gene expression profile in this study contrasting with 21d/40d ($P<0.01$) and 40d/70d ($P<0.05$) showed a single peak of high expression at 40 days of pregnancy (**Figure 2**), which coincide with the primary generation stage of muscle formation which is between 35 and 60 days of pregnancy (Wigmore and 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 encodes proteins from the MYBPC (myosin binding protein C) family which is a sarcomeric protein, whose role in sarcomere structure and regulation of contraction is currently under investigation (Oakley et al. 2007). The MYBPC1 gene is known to be the only MYBPC gene expressed in skeletal muscle prior to birth (Gautel et al., 1998 and Kurasawa et al., 1999). Its expression profile showed a low expression at the first period (21d/40d, $P<0.01$) 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 fiber formation which is between 54 to 90 days of pregnancy (Wigmore and 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 ($P<0.01$), being more expressed during prenatal, demonstrating that it is playing a role during muscle development. TPM (tropomyosin) is a dimeric, coiled coil protein which interacts with troponin and actin, regulating skeletal muscle contraction and stabilizing the actin filament. In skeletal muscle, three major isoforms, TPM1, TPM2, and TPM3, are expressed. However, the conformation and function of the TPM dimmer is obscure, since the identification and quantification of the isoforms have not been clarified in the past studies (Oe et al. 2007). The postnatal expression profile found in this work agrees with results found by Różycki et al. 2011, where the expression level of TPM2 did not change significantly.

The p53 is encoded by human gene known as TP53; it is a tumor suppressor protein and responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. This protein is expressed at low level in normal cells and at a high level in a variety of transformed cell lines, where it's believed to contribute to transformation and malignancy (May and May, 1999; Tyner et al., 2002). Here at the major network it is connected only with the subnet Wnt signaling pathway which is strongly related with embryogenesis path according with its expression profile. This gene showed to be highly expressed at the first period (21 days prenatal) going down at 40 days prenatal, and differing from all three postnatal periods ($P<0.01$).

The DCTN1 had a high expression at the first period (21d) but its expression at 40 days was also high ($P<0.05$) when comparing with the other five periods (**Table 3**).The DCTN1 gene provides a protein called dynactin-1 which has at least two different versions differing in size; the larger version is called p150-glued, and the smaller version is called p135. Dynactin contributes significantly to cytoplasmic organization and subcellular motility at all stages of the cell cycle. Its ability to bind motors, microtubules, and cargo structures allows dynactin to link a wide variety of cellular components to the cytoskeleton (Schroer, 2004). For these reasons it is believed to be connected with motor activity subnet. The DCTN1 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 (**Figure 2**). ANKRD2 is a member of the Muscle Ankyrin Repeat Protein family (MARPs), consisting of sarcomere-associated proteins that can also localize in the nucleus. Results in mice suggest that ANKRD2 gives an important contribution to the coordination of proliferation and apoptosis during myogenic differentiation in vitro, mainly through the p53 network (Bean et al. 2008). Another gene from this cluster, Nebulin (NEB) is a giant modular sarcomeric protein that has been proposed to play critical roles in myofibrillogenesis, thin filament length regulation, and muscle contraction (Bang et al. 2006). It is present at the same subnets as ANKRD2: myofibril, muscle system process and structural constituent of the muscle. In **Figure 2** we can see the ANKRD2 and NEB expression level being high at 21 days ($P<0.01$) and after that being low across all other six periods.

Differing on this cluster we have two genes, Obscurin (OBSCN) and Ankyrin 1 (ANK1). Both genes expression profile had a higher point at 21 days ($P<0.01$), being low 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). The OBSCN encodes a protein which belongs to the family of giant sarcomeric signaling proteins which includes Titin and Nebulin, and may have a role in the organization of myofibrils during assembly and may mediate interactions between the sarcoplasmic reticulum and myofibrils (Young et al. 2001). There are reports about this 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. The ANK1 gene belongs to a family of proteins that link the integral membrane proteins to the underlying spectrin-actin cytoskeleton and plays key roles in activities such as cell motility, activation, proliferation, contact and the maintenance of specialized membrane domains (Wimmers et al., 2007), and it was linked only with myofibril subnet.

Postnatal over expressed

On this cluster 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. The two first belongs to a sarcomeric myosin heavy chain (MHC) family consisting of at least eleven isoforms (Weiss et al., 1996). The MYH2, MYH1 and MYH4 genes encode the protein isoforms commonly termed MHC IIA, IIX and IIB (Tonge et al., 2010). Of the eleven sarcomeric isoforms genes of MHC, four are known to be expressed in adult skeletal muscle: one "slow-twitch" (Type I or MYH7) muscle-associated MHC isoform and three "fast-twitch" (Types IIA, IIX and IIB) muscle-associated isoforms. Here MYH2 and MYH7 are at the same subnets, also with Wnt signaling 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 statistically less expressed which was at 21d and 70d ($P<0.05$), while MYH7 gene had all prenatal periods being less expressed than postnatal ($P<0.01$).

As in MYH2 and MYH7, ACTA1 and TNNT3 genes showed a similar expression profile. ACTA1 is founded in six different actin isoforms which are highly conserved across species with more than 90% similarity in their amino acid sequence. Four of them are muscle-specific isoforms which are expressed in different types of muscle tissues, and are functionally involved in muscle contraction (Ilkovski et al., 2005); the α-skeletal actin (ACTA1) is a very important type in this group. It was linked to the Wnt signaling pathway, myofibril and muscle system process according to its property. Here its expression profile showed low levels at prenatal stage ($P<0.01$) and high at postnatal. Troponin T (TnT) in vertebrates occurs in three tissue-specific forms of striated muscle: fast skeletal TnT, slow skeletal TnT and cardiac TnT (Perry, 1998), each isoform is encoded by a separate gene, in humans TNNT3, TNNT1 and TNNT2, respectively (Mao et al., 1996). There are reports where the expression levels of TNNT3 gene showed to be not differentially related to the pig slaughter age of Polish commercial animals line (Różycki et al., 2011). In our finds this gene had a higher expression during postnatal stage while the expression level in prenatal stage was lower ($P<0.01$), 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 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 a high expression during fetal 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.

Besides, new candidates for muscle development (TP53 and DCTN1) can be listed. Here they showed a high expression level during the first muscle development wave which was at 21 and 40 days of pregnancy and are connected with Wnt signaling pathway and motor activity subnets respectively. The network constructed can be helpful to better understand the molecular mechanisms involved with skeletal muscle. However, this is a complex trait that is subject to action from a large number of genes in muscle that are regulated by several transcription factors.

Conclusion

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 differentially expressed genes mapped in the same genetic network. The present study also provides rich new information resource about genes from which are cited to be involved in myogenesis for the first time increasing our understanding of the molecular mechanisms underlying pig skeletal muscle development. Our comparative analysis of the prenatal and postnatal skeletal muscle specific genes suggests that there are genes which play important role for one stage more than to another. This finding could contribute to better explaining genes function mechanism. 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 differentially expressed genes were identified across stages associated with meat production traits, which may be commercially valuable.

Authors’ contributions

SEFG and PSL conceived of the study, and participated in its design and coordination. CSN and JOP performed of the biological analysis and the EST sequencing under the guidance of MFMG and SEFG. LLV and CSN were responsible for the bioinformatics and qRT-PCR analysis and wrote the manuscript with the help of SEFG. All authors have read and approved the final manuscript.

Acknowledgements

We would like to thank the Brazilian agencies FAPEMIG, CAPES, FINEP and CNPq for their financial support.

References

Altschul SF, Gish W, Miller W, Myers EW et al. **Basic local alignment search tool.** *Journal of Molecular Biology 1990*, **215**:403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang J et al. **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs**. *Nucleic Acids Research 1997*, **25**:3389–3402.

Bang ML, Li X, Littlefield R, Bremner S et al. **Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle.** *Journal of Cell Biology* 2006, **173**:905-916.

Barton P J, Buckingham M E: **The myosin alkali light chain proteins and their genes.** *Biochemical Journal* 1985, **231**: 249-261

Bean C, Facchinello N, Faulkner G, Lanfranchi G: **The effects of Ankrd2 alteration indicate its involvement in cell cycle regulation during muscle differentiation.** *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2008, **1783**(6):1023-1035.

Dittmer D P, Gonzalez C M, Vahrson W, DeWire S M, Hines-Boykin R and Damania B. **Whole-Genome Transcription Profiling of Rhesus Monkey Rhadinovirus.** *Journal of Virology,* 2005, 79 (13):8637-8650.

Eisen M B, Spellman P T, Brown P O and Botstein D. **Cluster analysis and display of genome-wide expression patterns**. *Proceedings of National Academy of Sciences of United States of America 1998* 95:14863–14868.

Gautel M, Furst DO, Cocco A, Schiaffino S: **Isoform transitions of the myosin binding protein C family in developing human and mouse muscles: lack of isoform transcomplementation in cardiac muscle.** *Circulation Research* 1998, **82**:124-129.

Gorodkin J., Cirera S., Hedegaard J. et al: **Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags.** *Genome Biology* 2007, **8:**R45

Greller LD, Tobin FL: **Detecting selective expression of genes and proteins**. *Genome Research* 1999, **9**: 282–296.

Hailstones D L, Gunning P W: **Characterization of human myosin light chains 1 sa and 3nm: implications for isoform evolution and function.** *Molecular and Cell Biology* 1990 **10**: 1095-1104.

Ilkovski B, Clement S, Sewry C, North KN, Cooper ST: **Defining alpha-skeletal and alpha-cardiac actin expression in human heart and skeletal muscle explains the absence of cardiac involvement in ACTA1 nemaline myopathy.** *Neuromuscular Disorders* 2005, **15**:829–835.

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

Livak KJ, Schmittigen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2DDCt method**. *Methods* 2001, **25**: 402-408.

Mao C., Baumgartner A.P., Jha P.K., Huang T.H.-M. Sarkar S: **Assignment of the human fast skeletal troponin T gene (TNNT3) to chromosome 11p15.5: evidence for the presence of 11pter in a monochromosome 9 somatic cell hybrid in NIGMS Mapping Panel 2.** *Genomics* 1996, **31**:385-388.

May, P. and May, E: **Twenty years of p53 research: structural and functional aspects of the p53 protein.** *Oncogene* 1999, **18:**7621–36.

Murray D, Doran P, MacMathuna P, Moss AC: ***In silico* gene expression analysis – an overview.** *Molecular Cancer* 2007, **6**:50.

Oakley CE, Chamoun J, Brown LJ, Hambly BD: **Myosin binding protein-C: enigmatic regulator of cardiac contraction.** *The International Journal of Biochemistry and Cell Biology* 2007, **39**:2161-2166.

Oe M, Ohnishi-Kameyama M, Nakajima I, Muroya S, Chikuni K: **Muscle type specific expression of tropomyosin isoforms in bovine skeletal muscles.** *Meat Science* 2007, **75**: 558–563.

Pfaffl MW: **A new mathematical model for relative quantification in real time RT-PCR**. *Nucleic Acids Research* 2001, **29**(9):e45.

Perry S.V: **Troponin T: genetics, properties and function.** *Journal of Muscle Research and Cell Motility* 1998, **19:**575-602.

Radoniæ A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A: **Guideline to reference gene selection for quantitative real-time PCR.** *Biochemical and Biophysical Research Communications* 2004, **313:**856-862.

Rehfeldt, C, I Fiedler, G Dietl, and K Ender: **Myogenesis and postnatal skeletal muscle cell growth as influenced by selection.** *Livestock Production Science* 2000, **66**:177–188.

Rehfeldt, C, I Fieldler and NC Stickland: **Number and Size of Muscle Fibres in Relation to Meat Production.** In: MFW te Pas, ME Everts and HP Haagsman, *Muscle development of livestock animals: physiology, genetics, and meat Quality* 2004, p. 15 – 52. CABI Publishing: Oxon.

Romualdi C, Bortoluzzi S, and Danieli GA. **Detecting differentially expressed genes in multiple tag sampling experiments. Comparative evaluation of statistical tests**. *Human Molecular Genetics* 2001, **10**: 2133–2141.

Różycki M, Piórkowska K, Oczkowicz M. **Expression of contractile protein-encoding genes TPM2 and TNNT3 during ontogenesis in pigs raised in Poland.** *Animal Science Papers and Reports,* 2011 **29** (1):65-74

Ryu, YC, Lee M-H, Lee S-K and Kim BC. **Effects of muscle mass and fiber type composition of *Longissimus dorsi* muscle on postmortem metabolic rate and meat quality.** *Journal of Muscle Foods* 2006 17: 343-352.

Schroer TA. **Dynactin.** *Annu. Rev. Cell Dev. Biol.* 2004, 20:759-79. Review.

Serão N, Veroneze R, Ribeiro A, Verardo LL, Braccini Neto J, Gasparino E, Campos C.F, Lopes P.S, Guimarães S.E.F: **Candidate gene expression and intramuscular fat content in pigs**. *Journal of Animal Breending and Genetics* 2010, JABG, **09:**0151.R2.

Sollero BP, Guimarães SEF, Rilington V D, Tempelman R J, Raney N E, Steibel J P, Guimarães J D, Lopes P S, Lopes M S and Ernst C W. **Transcriptional profiling during foetal skeletal muscle development of Piau and Yorkshire–Landrace cross-bred pigs**. *Animal Genetics,* doi:10.1111/j.1365-2052.2011.02186.x

Steibel JP, Poletto R, Coussens PM, Rosa GJM: **A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data**. *Genomics* 2009, **94:**146–152.

Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries.** *Genome Research* 2000, 10: 2055–2061.

te Pas MF, De Wit AA, Priem J, Cagnazzo M, Davoli R, Russo V, Pool MH: **Transcriptome expression profiles in prenatal pigs in relation to myogenesis.** *Journal of Muscle Research and Cell Motility* 2005, **26**:157-165.

Tonge DP, Jones SW, Bardsley RG, Parr T: **Characterisation of the sarcomeric myosin heavy chain multigene family in the laboratory guinea pig.** *BMC Molecular Biology,* 2010, **11:**52.

Tyner SD, Venkatachalam S, Choi J, Jones S, Ghebranious N, Igelmann H, Lu X, Soron G, Cooper B, Brayton C, Hee Park S, Thompson T, Karsenty G, Bradley A, Donehower LA: **p53 mutant mice that display early ageing-associated phenotypes**. *Nature* 2002, **415** (6867): 45–53.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes.** *Genome Biology* 2002, 3:research0034-research0034.11

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

Weiss A, Leinwand LA: **The mammalian myosin heavy chain gene family.** *Annual Review of Cell and Developmental Biology* 1996, **12:**417-439.

Wigmore PMC and Stickland NC: **Muscle development in large and small pig fetuses.** *Journal of Anatomy* 1983, **137**: 235–245.

Wimmers K, Murani E, TePas M. F, Chang K. C, Davoli R, Merks J. W, Henne H, Muraniova M, daCosta N, Harlizius B, Schellander K, Böll I, Braglia S, deWit A. A, Cagnazzo M, Fontanesi L, Prins D, Ponsuksili S. **Associations of functional candidate genes derived from gene-expression profiles of prenatal porcine muscle tissue with meat quality and muscle deposition.** *Animal Genetics* 2007, **38**:474-484.

Young P, Ehler E, Gautel M: **Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly.** *The Journal of Cell Biology* 2001 154 (1):123–36.

Zhang S-Z, Xu Y, Xie HQ, Li X-Q, Wei W-Q, Yang Z-M: **The possible role of Myosin light chain in myoblast proliferation.** *Biology Research.* 2009, **42**:121-132.

Supplementary data

Table S1. Deferentially expressed genes primer concentration and cDNA. Their concentrations were determinate through amplification efficiency as described by Pfaffl, 2001.

Genes*	Primer Concentration (nM)	cDNA concentration (ng)
ANKRD2	400	75
MYH2	400	25
ACTA1	200	25
DCTN1	400	25
MYBPC1	400	25
NEB	400	225
TTN	100	75
TPM2	400	75
OBSCN	100	25
ANK1	100	25
MYH7	400	75
MYL2	100	25
TNNT3	100	25
TP53	400	75

*ANKRD2- ankyrin repeat domain 2; MYH2- myosin, heavy chain 2; ACTA1- actin, alpha 1; DCTN1- dynactin 1; MYBPC1- myosin binding protein C; NEB- nebulin; TTN- titin; TPM2- tropomyosin, 2; OBSCN- obscurin; ANK1- ankyrin 1; MYH7- myosin, heavy chain 7; MYL2- myosin, light chain 2; TNNT3- troponin T, type 3; TP53- tumor protein, p53.

Table S2. Genes whose transcripts varied significantly depending on the abundance of ESTs in the libraries of three cDNA pig *Semi-membransous* muscle. 34 contigs with functional annotation showed differential expression. We applied the AC statistic, Fisher's 2x2 exact test and 2x2 Chi^2 test, and in multiple comparisons, the R statistic, GT statistic, and general Chi^2 test were used. Cells with different degrees of blue color represent extent of gene expression.

Gene Name	**Description**	**Duroc (norm)***	**LW (norm)***	**Piau (norm)***
Differentially expressed gene name in Large White				
MT-CYB	mitocondrially encoded cytochrome b	0	11.4	68.7
MyH1**	myosin, heavy chain 1, skeletal muscle, adult	361.4	928.7	215.8
MyH2*	myosin, heavy chain 2, skeletal muscle, adult	23.9	70.9	8.9
Differentially expressed gene name in Piau				
ANK1**	ankyrin 1, erythrocytic	35.8	0	0
MyPLF**	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform	18.3	0	154.5
NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	0	0	29.4
TNNT3**	troponin T type 3 (skeletal, fast)	47.8	0	0
TPT1	tumor protein, translationally-controlled 1	52.5	50.1	214.2
Differentially expressed gene name in Duroc				
ACTA1**	actin, alpha 1, skeletal muscle	888.9	697.7	434.8
ACY1	aminoacylase 1	23.9	0	0
ANKRD2**	ankyrin repeat domain 2 (stretch responsive muscle)	361.4	21.6	17.1
ATP5A1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1	23.9	0	0
ATP6V1H	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H	119.5	4.6	62.1
CA3	carbonic anhydrase III	444.4	232.6	86.1
CGGBP1	CGG triplet repeat binding protein 1	23.9	0	0
DCTN1**	dynactin 1	175.1	0	165.9
DYSFIP1	dysferlin interacting protein 1	35.8	2.3	0
ENO3	enolase 3 (beta, muscle)	888.9	620.2	173.9
EXOSC8	exosome component 8	47.8	0	0
MT-COX3	mitochondrially encoded cytochrome c oxidase III	71.7	38.9	173.9
MyBPC1**	myosin binding protein C, slow type	240.1	0	0
MyBPH**	Myosin-binding protein H (MyBP-H) (H-protein)	47.8	0	0
MyH7**	myosin, heavy chain 7, cardiac muscle, beta	481.9	21.6	17.1
MyL2**	slow cardiac myosin regulatory light	240.1	64.8	431.7
NAGA	N-acetylgalactosaminidase, alpha	23.9	0	0
NDUFA9	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	36.7	0	0
PDLIM5**	PDZ and LIM domain 5 isoform d	444.4	77.5	173.9
RPL10A	ribosomal protein L10a	23.9	0	0
SDHS	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	36.7	0	29.4
SMARCD3**	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3	23.9	0	0
TCBE2	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)	444.4	0	0
TPM2**	tropomyosin 2 (beta) isoform 1	1111.1	542.6	434.8
TRIM54**	tripartite motif-containing 54	361.4	0	17.1
ZCRB1	zinc finger CCHC-type and RNA binding motif 1	47.8	0	0

*Normalized number of EST in each Contig. **16 muscle-specific genes used for analyses.

Table S3. 36 genes showing differential expression. The results are shown for the three libraries (Duroc, Large White and Piau) and considering six different statistical tests with the p-values. The associated significance thresholds (<0.05) were applied using the Bonferroni correction. (To be continued).

Genes	**ID**	**AC (1 2)**	**AC (1 3)**	**AC (2 3)**	**Fisher (1 2)**	**Fisher (1 3)**	**Fisher (2 3)**	**Chi2x2 (1 2)**	**Chi2x2 (1 3)**	**Chi2x2(2 3)**	**GT**	**R**	**Chi**
DEGs in Large White breed													
MT-CYB	ENSP00000354554	0,000799	0,000004	1	0,010405	0,000013	0,327599	0,016147	0,000022	0,463357	0,000002	0,000009	11
MYH1	NM_005963.3	0,006753	0,02439	0	0,129446	0,428254	0	0,086639	0,413731	0,000001	0,366062	0,000003	0,000002
MYH2	NM_001100112.1	0,009594	0,029121	0	0,152401	0,238915	0,000001	0,116226	0,233071	0,000003	0,502402	0,000004	0,00001
DEGs in Piau breed													
ANK1	NM_000037.3	0,047144	0,000948	0,000053	0,606566	0,009992	0,000198	0,246641	0,01797	0,000221	0,398429	0,000023	0,0001
MYLPF	NM_013292.3	0,114874	0,001515	0,000002	0,33893	0,008287	0,000006	0,162407	0,011965	0,000047	0,638335	0,000001	0,000015
NDUFB2	NM_004546.2	.	0,074292	0,043505	.	0,583268	0,13592	.	0,204844	0,076678	.	0,043972	0,093514
TNNT3	NM_001001863	0,047144	0,000948	0,000053	0,606566	0,009992	0,000198	0,246641	0,01797	0,000221	0,398429	0,000023	0,0001
TPT1	NM_003295.2	0.056516	0.001276	0.000006	1.000.000	0.011107	0.000017	0.941064	0.011017	0.000012	0.338239	0.000021	0.000011
DEGs in Duroc breed													
ACTA1	NM_001100.3	0,055974	0,042533	0,079433	0,743305	0,27012	0,421443	0,674441	0,262335	0,378085	0,230367	0,522174	0,505246
ACY1	NM_000666.1	0,004151	0,003836	.	0,025803	0,024479	.	0,001228	0,001026	.	.	0,007411	0,000025
ANKRD2	NM_020349.2	0,001811	0,000991	0,225291	0,012117	0,00769	1	0,000828	0,000215	0,896774	1	0,010451	0,000049
ATP5A1	NM_001001937.1	0,004151	0,003836	.	0,025803	0,024479	.	0,001228	0,001026	.	.	0,007411	0,000025
ATP6V1H	NM_015941.2	0,000667	0,0006	.	0,004139	0,003824	.	0,000075	0,000058	.	.	0,000638	0
CA3	NM_005181.3	0,070487	0,047971	0,147288	0,605009	0,191127	0,6244	0,46382	0,133905	0,371287	0,184592	0,369561	0,349155
CGGBP1	NM_001008390.1	0,004151	0,003836	.	0,025803	0,024479	.	0,001228	0,001026	.	.	0,007411	0,000025
DCTN1	NM_001135040.1	0.000000	0.041658	0.000000	0.000000	0.849613	0.000000	0.000000	0.884114	0.000000	0.285030	0.000000	0.000000
DYSFIP1	NM_001007533.3	0,00224	0,0006	0,242236	0,014565	0,003824	0,492174	0,001326	0,000058	0,309709	0,330332	0,002719	0,000012
ENO3	NM_053013.2	0,052269	0,013637	0,032257	0,510259	0,053098	0,107951	0,540166	0,032332	0,079243	0,249073	0,094594	0,104456

Table S3. 36 genes showing differential expression. The results are shown for the three libraries (Duroc, Large White and Piau) and considering six different statistical tests with the p-values. The associated significance thresholds (<0.05) were applied using the Bonferroni correction. (Continuation)

Genes	ID	AC (1 2)	AC (1 3)	AC (2 3)	Fisher (1 2)	Fisher (1 3)	Fisher (2 3)	Chi2x2 (1 2)	Chi2x2 (1 3)	Chi2x2(2 3)	GT	R	Chi
DEGs in Duroc breed													
EXOSC8	NM_181503.2	0,000107	0,000094	.	0,000663	0,000597	.	0,000005	0,000003	.	.	0,000055	0
MT-COX3	ENST00000362079	0,052269	0,013637	0,032257	0,510259	0,053098	0,107951	0,540166	0,032332	0,079243	0,249073	0,094594	0,104456
MYBPC1	NM_002465.2	0,003513	0,002191	.	0,022872	0,016694	.	0,000819	0,000246	.	.	0,005667	0,000005
MYBPH	NM_001033014	0,000107	0,000094	.	0,000663	0,000597	.	0,000005	0,000003	.	.	0,000055	0
MYH7	NM_000257.2	0,000344	0,000161	0,225291	0,002216	0,001202	1	0,00005	0,000008	0,896774	1	0,001239	0
MYL2	NM_000432.3	0,057046	0,047971	0,208944	0,275222	0,191127	1	0,264682	0,133905	0,629998	0,153154	0,374318	0,290041
NAGA	NP_000253.1	0,004151	0,003836	.	0,025803	0,024479	.	0,001228	0,001026	.	.	0,007411	0,000025
NDUFA9	NM_005002.3	0,038934	0,023452	.	0,114734	0,081826	.	0,04812	0,025459	.	.	0,033707	0,011713
PDLIM5	NM_001011515.1	0,038473	0,068958	0,186267	0,164441	0,314527	0,602985	0,103493	0,324396	0,495206	0,153154	0,333694	0,26687
RPL10A	NM_007104.4	0,004151	0,003836	.	0,025803	0,024479	.	0,001228	0,001026	.	.	0,007411	0,000025
SDHA	NM_004168.2	0,038934	0,091304	0,043505	0,114734	1	0,13592	0,04812	0,798249	0,076678	0,246444	0,067531	0,177437
SMARCD3	NM_001003801.1	0,038934	0,023452	.	0,114734	0,081826	.	0,04812	0,025459	.	.	0,033707	0,011713
TCEB2	NM_207013.1	0,017298	0,022247	.	0,065776	0,07783	.	0,016026	0,022904	.	.	0,024221	0,004254
TPM2	NM_003289.3	0,029173	0,023925	0,112425	0,302072	0,145592	0,773391	0,195043	0,112058	0,697351	1	0,330643	0,251603
TRIM54	NM_032546.2	0,000534	0,000991	0,247918	0,003406	0,00769	1	0,000041	0,000215	0,361247	0,330332	0,00204	0,000002
ZCRB1	NM_033114.3	0,000107	0,000094	.	0,000663	0,000597	.	0,000005	0,000003	.	.	0,000055	0

Chapter 3

NON-CODING RNAs IDENTIFICATION AND EXPRESSION LEVELS OF PIG miRNAs IN SKELETAL MUSCLE

Abstract. MicroRNAs are a class of naturally occurring non-coding RNAs. Typically they have ~22 nucleotides (nt) in length that suppress translation of their targets genes. Several laboratories have attempted to identify miRNAs from pigs and the bioinformatics strategies using ESTs have proved to be successful for this aim. In this study we report an *in silico* identification of ncRNA in pig EST libraries focusing on novel pig miRNAs and further investigated the differential expression of pigs miRNAs (known and novel) by quantitative real-time PCR during pre- and postnatal stage from commercial pig breed skeletal muscle tissue. From a total of 2286 EST sequences it was possible to identify by computational approaches two miRNAs showing an appropriate secondary structure and being not yet described in pigs. A substantial number of remaining no hit sequences showed homology with non-coding sequences as a Piwi-interacting RNA (piRNA) and others. Predicting the targets of miRNAs which had homology with pigs' sequences we found a total of 503 potential miRNA targets for them. The miRNA which had more number of targets was the hsa-miR-1207-5p, with 288 targets followed by hsa-miR-665 with 214. From these targets, four were muscle specific genes being three for hsa-miR-1207-5p (CAMK2D, AK1 and SLC25A34), and just one target identified for the hsa-miR-665 (DIDO1). The protein-protein interaction was determined with these four genes finding nine related genes. They formed a specific network composed for subnets like kinase activity, programmed cell death, synaptic transmission, cell projection and genes involved in the activation of NMDA receptor upon glutamate binding and postsynaptic events. The expression analyses were performed for known and novel pigs miRNA across seven muscle developmental periods in which a differential between pre- and postnatal stages was observed. This work suggests that EST analysis is a feasible strategy for identifying ncRNAs as new miRNA candidates with their targets. The findings of miRNAs and their muscle-specific targets in pigs will be helpful for understanding the function and processing of pig small RNAs in the future. Besides, the expression level of the miRNAs analysed across pre- and postnatal stages demonstrate a role on biological events during muscle development and may be worthy of further investigation on their biological role in pigs.

Keywords EST, skeletal muscle, miRNA, targets, expression

Introduction

MicroRNAs are a class of naturally occurring non-coding RNAs (ncRNA), which are expressed in a wide range of eukaryotic organisms (Bartel et al., 2004). Typically they have ~22 nucleotides (nt) in length that suppress translation of their targets genes by binding to the 3' untranslated region (Ambros, 2004 and Bartel, 2004). Several laboratories have attempted to identify miRNAs from pig (Huang et al., 2008) and the bioinformatics strategies have proved to be successful for this aim. This approach is based on genome sequences or other databases such as expressed sequence tags (ESTs) and genome survey sequences (GSS) which are nucleotide sequences similar to EST's, with the exception that most of them are genomic in origin, rather than mRNA(Xie et al., 2010 and Zhang et al., 2006). Previous research has used the public available ESTs to search new miRNA genes in pigs using the previously known miRNAs from human and mouse and others mammals non-coding sequences (Zhou & Liu, 2010 and Seeman et al., 2007).

Recent studies have shown that miRNAs play important gene-regulatory roles in numerous eukaryotic lineages and are often highly conserved across animal species. They are involved in many diverse biological processes and may potentially regulate the functions of thousands of genes. The regulation mechanisms can be since a repression of translation to a cleavage of targeted mRNAs. Also, they may directly target transcription factors which affect animal development and specific genes which control metabolism (Carthew, 2006). Identification of comprehensive sets of miRNAs and other small regulatory RNAs in different organisms is a critical step to facilitate our understanding of genome organization, genome biology and evolution (Carrington & Ambros, 2003), as the different stages of muscle development.

Skeletal muscle development is an important physiological process in animal production, and it directly affects meat production. Muscle mass is mainly determined by muscle fiber number and size in animals. In pigs, muscle fibers are formed in two stages during gestation, including primary and secondary fiber formation, being the muscle fiber numbers fixed before birth. Investigation of genes expressed during skeletal muscle development is elementary in understanding molecular mechanism of muscle growth as the identification of miRNAs related with muscle specific genes. In this study we report a computational identification of non-coding sequences in owner EST libraries data base focusing in new pig miRNA with their appropriated secondary structure and further investigated differential expression of miRNAs by quantitative real time PCR during pre- and postnatal stage from skeletal muscle tissue in commercial pigs.

Material and Methods

All methods involving animal handling was done 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).

miRNAs dataset

To search potential miRNAs, a total of 2625 previously known miRNAs from *Homo sapiens* (human) and *Mus musculus* (mouse) were obtained from the miRNA Registry Database (Release 16, September 2010; http:/www.mirbase.org; Griffiths-Jones, 2004). The *Sus scrofa* (pig) miRNA dataset was provided by Li et al. (2010) and a total of 777 pig miRNA sequences were obtained. All theses sequences were used to construct a mature miRNA reference dataset to predict known and novel miRNA sequences on the available pig EST library. In addition, 510,055 ncRNAs sequences were downloaded from fRNAdb, a comprehensive functional RNA database freely available (http://www.ncrna.org/frnadb/). These sequences were used to predict other RNA classes.

EST dataset

All ESTs used in this study were generated from three different pig breeds *semimembranosus* muscle tissue cDNA libraries (Duroc, Large White and a naturalized Brazilian breed, Piau) (Nascimento et al., 2011) being 392, 1312 and 1438 ESTs sequences respectively. Each EST library was assembled in Contigs and Singlets by CAP3 Sequence Assembly Program (http://deepc2.psi.iastate.edu/aat/cap/cap.html) web server, (Huang & Madan, 1999) generating 150, 475 and 602 unique sequences in each dataset, respectively. The parameters used on this step were the default.

Besides, we kept short sequences with less than a hundred nucleotides of length for further analysis; being 145, 349 and 557 short sequences for Duroc, Large White and Piau pig breed respectively. Unique sequences originated by CAP3 program were used for alignment by BlastX program (2.2.18), freely available at (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastNews#1), against UniProtKB/Swiss-Prot protein database (http://www.ebi.ac.uk/swissprot/) using an E-value 10 to pick up all no hit sequences. All BlastX no hit and short sequences, totalizing 2278, were recorded as a distinct dataset and used on BlastN searches to predict known and novels miRNA.

Homology searches and secondary structure predictions

The basis for computational identification is the conserved mature miRNA sequence coupled with the predictable secondary structure for its primary miRNA transcript (Ambros et al., 2003; Sunkar & Jagadeeswaran, 2008; Zhang et al., 2006). Conserved miRNAs are highly identical in sequence among diverse animal species. We used the mature miRNA reference data set obtained from Human, Mice and Pigs as the subject in Blast searches against pig EST sequences including non-coding sequences and short EST sequences previously filtered out before the BlastX step. The BlastN algorithm was used to find homologous miRNA sequences for pig that matched for possible sequence variations by the following parameters: -W 4, –S 1, F F –q -4 –r 5 –b 5 and E-value 1, where –W is the word size, -S 1 is for forward direction, -F F is no filter, -q is the mismatch penalty, -r is the match score and –b is the alignment number.

Hits among the pig EST sequences and known miRNA with less than four nucleotides of mismatch between them were filtered out.

The sequences which did not have homology with pig miRNA but had with humans and mice were picked up to be mapped into a pig genome. For this, we used the MapMi web server available in http://www.ebi.ac.uk/enright-srv/MapMi/ (Guerra-Assunção & Enright, 2010). Eventually some could not be mapped, and then we carried out a BlastN of EST sequences against the pig genome database (ftp://ftp.ensembl.org/pub/current/fasta/sus_scrofa/dna/) to determine and confirm their locations into the genome. The BlastN parameters used to map miRNA sequences on pig genome, were: -G 1 -E 2 -W 15 -F "m D" –U and E-value 1e-20, where –G is the cost to open a gap, -E is the cost to extend a gap, -W is the word size and -F "mD" –U was used to mask during the seed generation stage, but with no sequence masked in the match extension stage. EST sequences that matched genomic sequences were selected and the flanking ~60 nucleotides were excised. On this, folding was performed using the program mfold version 3.5 (Markham & Zuker, 2005 and 2008) available at (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) web server.

Furthermore miRNA precursor sequences were considered a potential miRNA when fitting the following criteria: (1) the predicted mature miRNA had no more than four nucleotide substitutions when aligned with known *H. sapiens and M. musculus* mature miRNAs; (2) the RNA sequence can fold into an appropriate stem-loop hairpin secondary structure; (3) a mature miRNA sequence site is present in one arm of the hairpin structure and (4) the predicted secondary structure has a negative minimum free folding energy (MFE; ΔG kcal/mol) and 30–70% Adenine+Uracil content (Zhang et al. 2006). These criteria reduced false positives and required that the predicted miRNAs fit the criteria proposed by Ambros et al., 2003. Besides, we submitted the new miRNA precursor sequences to an *ab initio* based program called MiPred (Jiang et al. 2007) which identify pre-miRNA-like hairpin sequences classifying them in real or pseudo pre-miRNA. The structures that qualified for the succeeding steps were those which were classified as a real pre-miRNA-like hairpin by MiPred program and fully satisfied the criteria proposed by Zhang et al. (2006).

We also employed Blast searches to obtained pig sequences similar to ncRNAs found in the fRNAdb database. The BlastN parameters were as following: -r 5 –q -4 –W 7 –G 10 –E 6 and E-value 1, where –r is the match score, -q is the mismatch penalty, –W is the word size, –G is the cost to open a gap and -E is the cost to extend a gap. The steps in homology search used for identification in this study are shown in **Figure 1**.

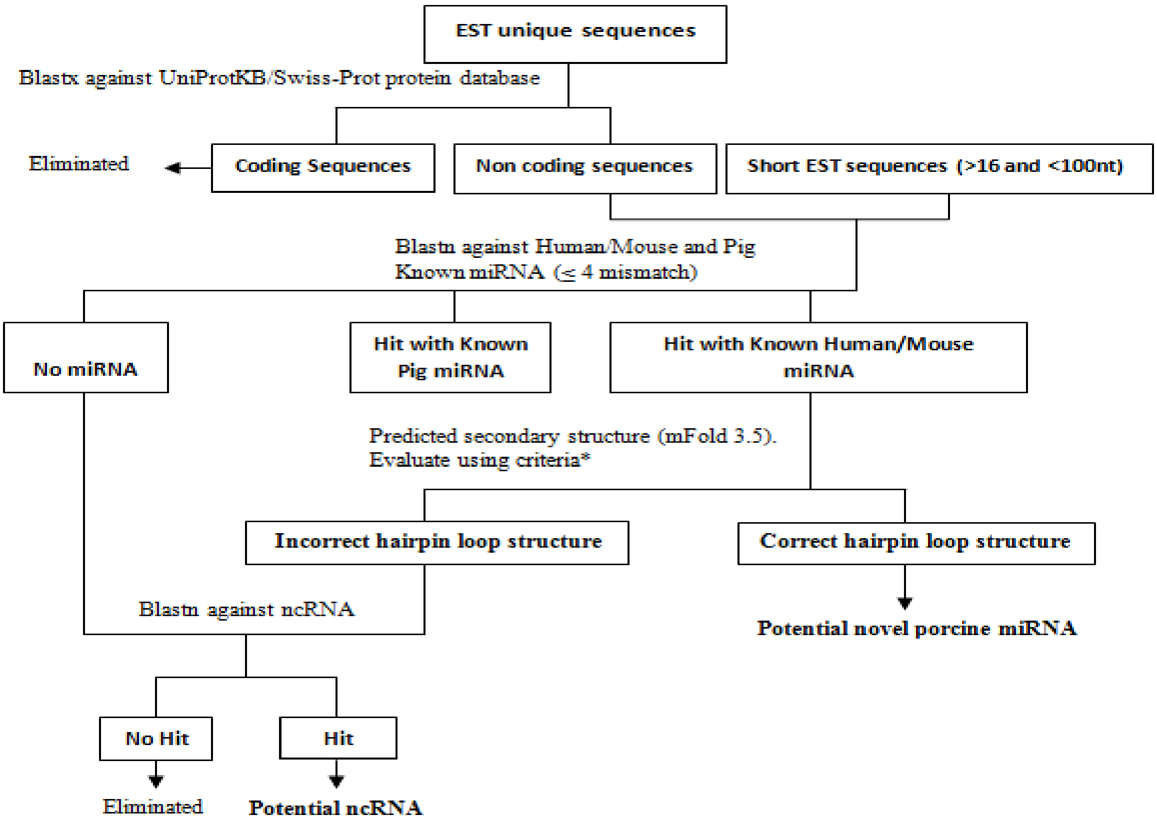


Figure 1. A schematic diagram of the over-all procedure used to identify and analyze miRNA and others ncRNAs. * modified from Zhang et al (2006), see text for more details.

Phylogenetic analyses

Considering the significance of miRNAs in evolution investigation, the precursor sequences of the predicted pig miRNAs and the known miRNAs in the same family were aligned and phylogenetically analyzed by ClustalW online with defaults parameters to investigate their evolutionary relationships (http://www.clustal.org/).

Potential miRNA target prediction

In the present study, new miRNAs targets were predicted using the web-based computational software TargetScan, which is publicly available at http://www.targetscan.org. *S. scrofa* genes are not included in the current version of TargetScan and the predictions were therefore based on the human mRNA/miRNA interactions. The identified known miRNAs targets were filtered against the TiGER muscle specific genes database available at http://bioinfo.wilmer.jhu.edu/tiger/db_tissue/muscle-index.html. This particular step served to provide information on potential targets of the putative miRNAs since sequence similarity with known mature miRNAs exists. In principle, since miRNAs have evolved to simultaneously regulate their targets, the miRNA targets would likely belong to a network having a consistent biological function. Thus, a gene network analysis was carried out using the miRNA muscle specific targets.

Gene Network Analysis

Aiming to examine the process of shared pathways, the Ensembl gene identifiers were used. These gene identifiers were obtained from target prediction and others proteins which interact to each other. The programs STRING (Szklarczyk et al. 2011) and TOPPCLUSTER [http://toppcluster.cchmc.org/] were used to obtain the protein-protein interaction and functional Gene Ontology (GO) respectively, identifying the biological mechanisms, pathways and functions involving these genes. The application Cytoscape [www.cytoscape.org/] was used to visualize and edit the pathways identified. By this way we could direct relationships between genes and their roles in muscle.

Quantitative real-time PCR of mature miRNA

Quantitative real-time PCR approach has been massively used on miRNA expression analysis (Chen et al. 2008; Huang et al. 2008 and Li et al. 2010). Here, it was used to measure the expression level pattern of miRNA genes during seven different periods of muscle development as described below.

Tissues samples

A total of 21 *Longissimus dorsi* (LD) muscle samples divided in both, pre- and postnatal stages were collected. At the UFV Pig Breeding Farm, abort was induced in pregnant Commercial gilts at 21, 40, 70 and 90 days of pregnancy using the following protocol: intramuscular injections of 1 mL Prelobam® (PGAF-α)-plus 1 mL Estrogen, followed 12 h later by 2 mL Orastina® (Ocitocine). LD muscle samples were isolated from 12 fetuses, 21d (n=3), 40d (n=3), 70d (n=3) and 90d (n=3); and placed in sterile tubes containing RNA*later*® (Qiagen). Samples were stored at 4ºC overnight and at -70ºC prior to RNA isolation. At the same way, samples of LD from Commercial castrated males in three different ages, 107 days (n=3), 121 days (n=3) and 171days (n=3) postnatal were isolated and stored. These time points cover major morphological and physiological changes of pig growth and development throughout pregnancy and up to 180 days after birth when the pigs reach peak commercial value.

RNA isolation and Reverse transcription

Total RNA from approximately 30-50 mg RNA*later*®-stabilized LD tissue from pig samples was isolated using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. RNA concentration was determined by analyzing 1 µl of solution using the ND-1000 micro-spectrophotometer (NanoDrop Technologies, Wilmington, DE). RNA integrity was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA integrity number (RIN) was determined using the RIN algorithm of the Agilent 2100 expert software. RNA with a RIN ≥ 5 was included in the study. cDNA was synthesized from 1 µg of total RNA using gene universal primer using the miScript Reverse Transcription (RT) Kit (Qiagen GmbH, Hilden, Germany), according to the manufacturer's instructions.

Quantitative Real-time PCR analysis

The qRT-PCR was performed in thermal cycler ABI Prism 7300 Sequence Detection Systems (Applied Biosystems, Foster City, CA, USA) using miScript SYBR Green PCR Kit (Qiagen) and amplification conditions for all systems were: 95 °C for 2 minutes, 40 cycles of denaturation at 95 °C for 15 seconds, and extension at 60 °C for 60 seconds. After 40 cycles of amplification, an additional step with a gradual increase in temperature of 60 to 95 °C was used to obtain the dissociation curve. The best primers and cDNA amplification conditions were used for qRT-PCR analyses. The amplification of target genes was performed at different wells and in duplicates (Pfaffl, 2001; Livak and Schmittigen, 2001). The primers used were published elsewhere (Chen et al., 2008), and constructed from miRNAs identified here and by Zhou et al., 2010 as novel in pigs. To make things easier they are listed in **Table 1**.

Table 1. Primers used for qRT-PCR reactions suplementar, amlpicon falar da temperature de dissociação

Gene	Access ID	Primer sequence (5'-3')	Reference
miR-424	MIMAT0001341	CAGCAGCAATTCATGTTTTGAA	Chen et. al. 2008
miR-133a	MIMAT0000145	TTTGGTCCCCTTCAACCAGCTG	Chen et. al. 2008
hsa-miR-1291	MIMAT0005881	TGGCCCTGACTGAAGACCAGCAG	Zhou, et. al. 2010
hsa-miR-147	MIMAT0000251	GTGTGTGGAAATGCTTCTG	Zhou, et. al. 2010
hsa-miR-1207-5p	MIMAT0005871	TGGCAGGGAGGCTGGGAGGG	testing
hsa-miR-665	MIMAT0004952	ACCAGGAGGCTGAGGCCCCT	testing

The experimental design used was completely randomized, with three replicates per period (21, 40, 70 and 90 days of pregnancy; 107, 121 and 171 days of adult age). Initially, data from qRT-PCR were analyzed using 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} ~N(0, σ^2_D); and E_{ijk} is a residual term., e_{ijk} ~ N(0, σ^2_e).

All statistical procedures were performed using SAS 9.0 for Windows (Statistical Analysis System Institute, Inc., Carry, NC, USA). The routine QPCR_MIXED: SAS® [https://www.msu.edu/~steibelj/JP files/QPCR.html] developed to generate commands in SAS PROC MIXED used to analyze qRT-PCR data was used in the analysis (Steibel et al., 2009). For each target gene, the comparison of expression values from a period to the following were performed by CONTRAST statement of the GLM procedure (SAS software) using t-student test.

Clustering relative quantification of gene expression

Hierarchical clustering was performed on PCR data (Eisen et al. 1998). ΔCt values (target Ct – endogenous Ct) were used in the analysis (Dittmer et al. 2005). The set of four miRNA genes expression values across seven periods was used for unsupervised hierarchical clustering based on Complete Linkage method with Pearson Correlation as a distance (Eisen et al. 1998).

Results

Homology searches and secondary structure predictions

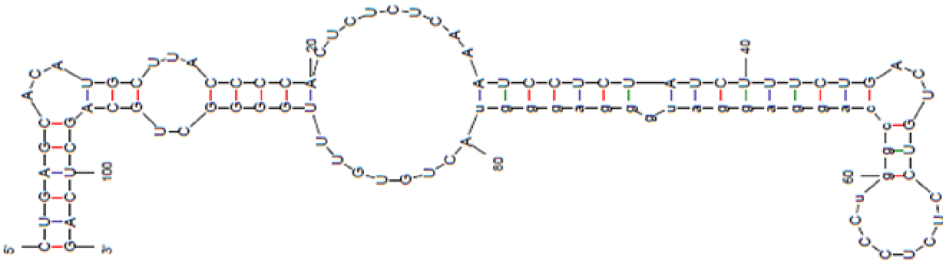
The BlastX against UniProtKB/Swiss-Prot protein database was used to filter out known coding sequences; 37, 88 and 109 no hit sequences were obtained for Duroc, Large White and Piau breeds dataset, respectively. They were blasted as well as the short EST sequences, against human, mouse and pig miRNA mature sequences using Blastn. After annotation, following the criteria cited in material and methods, only those sequences which had their predicted mature sequence with no more than four mismatches against mature miRNA sequence were selected. Sequences which had homology only with human and/or mouse were selected with only 14 sequences left. The mapping of these new miRNA was performed using their predicted mature sequence by MapMi tool. We found location for 12 sequences which were confirmed by BLAST against the pig genome, except for the related EST sequence Pi_Uf_74A08 which had discrepancies between Blast and MapMi genome location prediction so it was considered the Blast location. The short EST sequence Lw_Uf_96H03, which did not have match by Blast, was considered the MapMi location. All fourteen EST's mature miRNA locations can be verified in the supplementary data on **Table S1.**

Mature miRNA flanking sequences were analysed using the mfold web server to evaluate the ability to form thermodynamically stable hairpin structures following the previous criteria to be classified as a potential novel miRNA. The sequences were folded in a secondary structure with a low minimum free energy and they were submitted to MiPred. After secondary structures analyses (**Figure 1**) following the cited criteria, two novel potential miRNA with a real precursor classified by MiPred were confirmed. Their mature sequence had homology with hsa-miR-1207-5p and hsa-miR-665 hereafter identified as ssc-miR-new1 and ssc-miR-new2. Their mature sequences alignment can be verified at **Table 2**. The secondary structures of those two novel miRNA can be viewed at **Figure 2**.

ssc-miR-new1

Similar to hsa-miR-1207-5p

```
|     ACA   UUA     CUCUCUCAA       -         ACU   CUC
CUGAGC   UGC   CCCCA         AUCCUCU AUCUUUCUG   GUC   U
GACUCG   ACG   GGGGU         UGGGAGG UAGGAGGAC   CGG   C
^     ---   UC-     UUGUGUCA-       G         ---   UCC
```



ssc-miR-new2

Similar to hsa-miR-665

```
CGG---------|     A    GAA     G      C--       GGA
            ACCAGC GGCU   GCCCG UCCUAC   UCAGCGA   G
            UGGUCG UCGG   CGGGU GGGGUG   GGUCGCU   A
GGCGUCGGUGGG^     G    A--     G      AGU       AGU
```

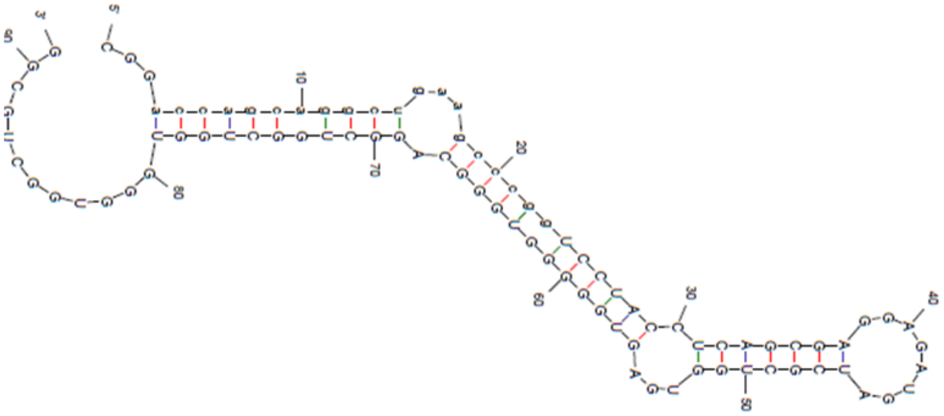


Figure 2. Hairpin secondary structures predicted by mFold 3.5 for the novel putative miRNAs with the mature miRNA sequences in red.

Table 2. Alignments of novel putative miRNAs with known miRNAs and their precursor's thermodynamic properties. Regions of mismatch are in red.

miRNA	EST source*	Aligned Candidate miRNA sequence (top) against Known miRNA (bottom).	MFE (ΔGKcal/ mol)	%A+U content
ssc-miR-new1	Pi_Uf_84A01	UGGCCAGGAGGAUGGGAGGGU (miR-new1) UGGCAGGGAGGCUGGGAGGGG (hsa-miR-1207-5p)	- 40.30	44,66
ssc-miR-new2	Lw_Uf_88G07/ Lw_Uf_80F11	ACCAGCAGGCUGAAGCCCGG (miR-new2) ACCAGGAGGCUGAGGCCCCU (hsa-miR-665)	- 44.70	34.78

*cDNA from sequence which the miRNA was identified, Pi and Lw means that these sequences are from Brazilian naturalized Piau and Large White breed libraries respectively.

A substantial number of remaining no hit sequences showed homology with non-coding sequences, after filtered out those which had homology with pig miRNA by Blast performance. A total of 1231 non-coding sequences were Blasted having 486 (39.5%) similarities with known ncRNA sequence as we can see on **Table 3**.

Table 3. Classes of others ncRNA identified.

RNA classes	Number of hits for each RNA class				
	Duroc	LW	Piau	Total	Percentage (%)
piRNA	23	71	92	186	38,3
Unclassified	12	61	87	160	32,9
rRNA	3	14	18	35	7,2
Intron	1	12	11	24	4,9
tRNA	3	7	12	22	4,5
3' element	1	8	5	14	2,9
IRES	2	1	4	7	1,4
snoRNA	2	4	0	6	1,2
snRNA	4	0	2	6	1,2
snmRNA	0	5	0	5	1,0
GSL3	3	2	0	5	1,0
TAR	0	2	2	4	0,8
RVS	0	1	2	3	0,6
Leader	0	1	1	2	0,4
3'UTR stem loop	0	0	2	2	0,4
PBS	0	1	1	2	0,4
Chanel RNA	0	1	0	1	0,2
pRNA	0	0	1	1	0,2
psi packing	0	0	1	1	0,2
Total number of sequences	**54**	**191**	**241**	**486**	**100**

PiRNA, piwi RNA; rRNA, RNA ribosomal; tRNA, transfer RNA; IRES, internal ribosome entry site; snoRNA, small nucleolar RNA; snRNA, small non-coding RNA; GSL3, gap stem loop 3; TAR, trans-activing region; RVS, Rev response element; PBS, protein biding site; pRNA, bacteriophage RNA.

Phylogenetic analyses

miRNA sequence comparisons of the members in the same miRNA families showed that the two mature miRNAs of the pig had a high degree of sequence similarity with the other members, but most of the precursor sequences were on the contrary, with sequence similarities score lower than 50 (**Table 4**) performed by ClustalW program. Based on the results of sequence comparisons, the precursor sequences of the members in mir-1207 and mir-665 families were phylogenetically analysed and the evolutionary relationship of the pig sequences was considered somewhat remote.

Table 4. Comparisons of sequence similarity between the pig miRNAs and corresponding miRNA family members, the similarity values are given by ClustalW program score.

		Similarity	
EST source*	**miRNA**	**Mature sequence**	**Precursor family**
Pi_Uf_84A01	hsa-miR-1207-5p	80.0	41.0
Lw_Uf_88G07/Lw_Uf_80F11	hsa-miR-665	80.0	38.6

*cDNA sequences which the miRNA was identified, Pi and Lw means that these sequences are from Brazilian naturalized Piau and Large White breed libraries respectively.

Potential miRNA target prediction

To identify the predicted targets related with miRNA identified in this work, we used the web server TargetScan. A total of 503 potential miRNA targets was found for the two miRNA and both revealed target multiplicity. The miRNA which has more number of targets is the hsa-miR-1207-5p with 289, while hsa-miR-665 has 214 targets. All targets were analysed seeking for those related to skeletal muscle, comparing them with muscle specific genes. Four muscle specific targets were identified. For miRNA hsa-miR-1207-5p three genes were identified (CAMK2D, AK1 and SLC25A34) and just one target was identified for the hsa-miR-665 (DIDO1).

Gene Networks Analyses

To understand the functions of those four targets genes and their protein interaction, we collected information about their biological process, cellular component and molecular function in the Gene Ontology. First, proteins which interact with miRNA targets were extracted from the STRING program, furthermore, using the application TOPPCLUSTER, we were able to identify their metabolic pathways. Thus, it was possible to identify 13 genes with obvious roles in muscle physiology allowing an insertion of 12 into a relevant functional metabolic network; the SLC25A34 target gene was not able to be inserted on it (see **Figure 3**). In addition, this major network is composed by five secondary networks connected with each other which are: kinase activity, programmed cell death, cell projection, synaptic transmission and genes involved in activation of NMDA receptor upon glutamate binding and postsynaptic events pathway. **Table 5** describes the function of the 12 genes which are present on the network plus SLC25A34 gene using the assignment of GO terms.

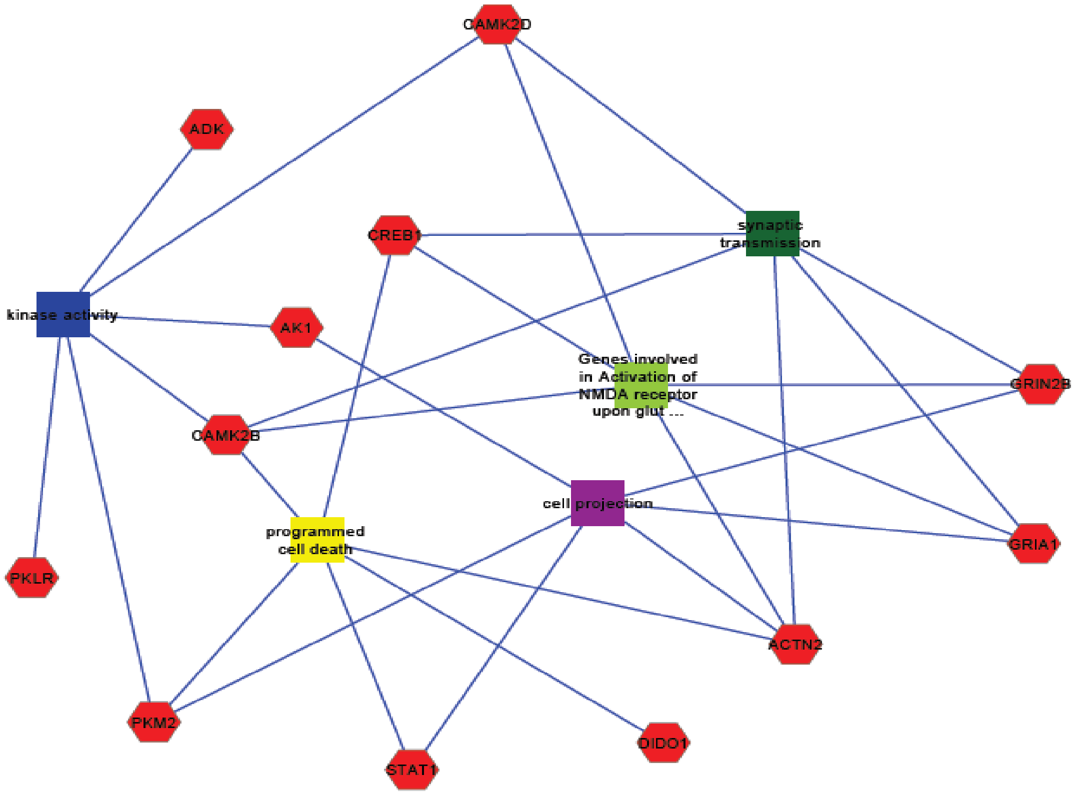


Figure 3. Functional gene networks and their interactions. It describes the relationships between three targets genes (AK1, CAMK2D and DIDO1) with nine linked genes (in red) and five important subnets; kinase activity (in blue), programmed cell death (in yellow), synaptic transmission (in dark green), cell projection (in purple) and genes involved in activation of NMDA receptor upon glutamate binding and postsynaptic events (in light green).

Table 5. GO mechanism of muscle specific targets of novel miRNAs and other interacting proteins.

Gene	GO: Molecular Function	GO: Biological Process	GO: Cellular Component	Pathway
GRIA1	receptor activity	synaptic transmission	cell projection	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
ADK	kinase activity	purine base metabolic process	nucleus	Purine metabolism
STAT1	DNA binding	transcription/programmed cell death	cell projection	Chemokine signaling pathway
AK1[a]	kinase activity	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	cell projection	Purine metabolism
CAMK2D[b]	kinase activity	regulation of cell growth	nucleus	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
PKM2	kinase activity	Glycolysis / Gluconeogenesis/programmed cell death	cell projection	carbohydrate metabolic process
ACTN2	actin binding	synaptic transmission/programmed cell death	cell projection	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
CAMK2B	kinase activity	synaptic transmission/programmed cell death	nucleoplasm	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
GRIN2B	receptor activity	synaptic transmission	cell projection	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
PKLR	kinase activity	response to hypoxia	soluble fraction	Glycolysis / Gluconeogenesis
CREB1	DNA binding	synaptic transmission/programmed cell death	chromatin	Genes involved in Activation of NMDA receptor upon glutamate binding and postsynaptic events
SLC25A34[c]	Binding	transport	mitochondrion	-
DIDO1[d]	protein biding	programmed cell death	nucleus	-

a, **b** and **c** are hsa-miR-1207-5p targets and **d** is hsa-miR-665 target

Quantitative real-time PCR of mature miRNA

The efficiencies of amplification by qRT-PCR represented around 100% amplification efficiency in each cycle, and the relative abundance was calculated using an equation to correct differences in efficiency as described by Pfaffl et al., 2001. Housekeeping gene such as the HPRT1, commonly used as internal controls for such analysis, were not suitable for normalization in these experiments because its transcription was altered across the periods. The hsa-miR-1291, which was consistently expressed in our study according to geNorm program (Vandesompele et al. 2002)**,** was therefore used as an internal control. The concentrations of primers and cDNA used in amplification reactions are shown in **Table S2**. We analysed the expression level of three known miRNAs (miR-424, miR-133a and hsa-miR-147) and two novels in pigs (hsa-miR-1207-5p and hsa-miR-665) across seven life time periods. From the two novel miRNAs identified in this work, only the hsa-miR-665 was amplified correctly and could be validated in pigs. Unsupervised hierarchical clustering based on the gene expression values (ΔCt) across all seven periods showed the genes expression levels (displayed as a heat map in **Figure 6**). The probability values for each contrast (gene/period versus period) demonstrate the significance of each contrast being significant and high significant p values < 0.05 and < 0.01, respectively in **Table 6**.

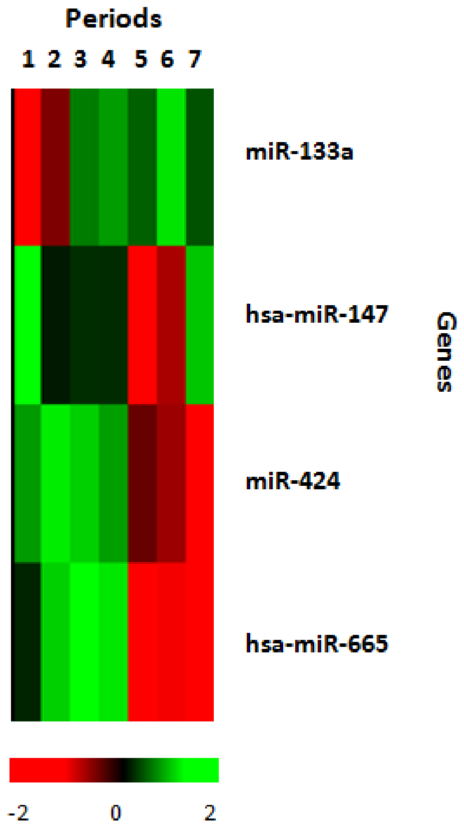


Figure 6. Unsupervised hierarchical clustering of miRNAs expression levels across all seven periods (1-4 prenatal, 21, 40, 70 and 90 days of pregnancy respectively and 5-7 postnatal, 107, 121 and 171 days after birth). They showed similar expression being a median expression value equal to zero was designated in black; green increased expression and red reduced expression.

Table 6. Differential expression p values for each miRNA in the comparison between periods.

	miRNA Genes			
Contrasts	**miR-424[1]**	**miR-133a[2]**	**hsa-miR-147[3]**	**hsa-miR-665[4]**
21d/40d	0.6593	<.0001**	0.1816	0.0045**
21d/70d	0.7774	<.0001**	0.2085	<.0001**
21d/90d	0.9838	<.0001**	0.2047	0.0018**
21d/107d	0.1858	<.0001**	0.0024**	<.0001**
21d/121d	0.1168	<.0001**	0.0417*	<.0001**
21d/171d	0.0032**	<.0001**	0.5237	<.0001**
40d/70d	0.8735	0.0013**	0.9327	0.0225*
40d/90d	0.6740	0.0005**	0.9419	0.6384
40d/107d	0.0868	0.0032**	0.0438*	<.0001**
40d/121d	0.0522	<.0001**	0.4273	<.0001**
40d/171d	0.0013**	0.0050**	0.4678	<.0001**
70d/90d	0.7929	0.6101	0.9908	0.0561
70d/107d	0.1152	0.6517	0.0370*	<.0001**
70d/121d	0.0703	0.1128	0.3813	<.0001**
70d/171d	0.0018**	0.5037	0.5199	<.0001**
90d/107d	0.1798	0.3423	0.0379*	<.0001**
90d/121d	0.1128	0.2614	0.3873	<.0001**
90d/171d	0.0031**	0.2471	0.5126	<.0001**
107d/121d	0.7836	0.0492*	0.1892	0.6637
107d/171d	0.0489*	0.8252	0.0097**	0.4190
121d/171d	0.0815	0.0322*	0.1384	0.2224

*$p < 0.05$ and ** $p < 0.01$ statistical differences between periods. miRNAs **1**, **2** and **3** are described in literature and **4** was identified in pigs on this work.

Discussion

We sought to identify novel miRNA genes across the unclassified sequences in our cDNA libraries which showed to be a useful source for this aim. The property of miRNAs to be highly conserved across closely related species was used in order to predict novel pig miRNA candidates (Kim et al., 2006). A Blast pipeline was performed suggesting the discovery of novel miRNA by computational approaches following the criteria proposed by Ambros et al., 2003. Though miRNAs are highly conserved across distantly related mammal species, in terms of primary and mature miRNAs (Berezikov et al., 2005), variations in pre-miRNA sequences provide the chance to investigate their evolutionary relationships. The two identified miRNA showed their mature sequence similar to hsa-miR-1207-5p and hsa-miR-665 but their precursor sequences were not, when compared with their putative miRNA families. That is, the precursor sequences of the pig miRNAs identified here are poorly conserved, according to studies that showed the occurrence of non conserved miRNAs and their precursors in mammals (Bentwich et al. 2005 and Li et al. 2010). It could be seen on **Table 3** that the two new pig miRNA precursor sequences were somewhat remotely related to others of the same miRNA family, comparing with the studies from pigs and other mammals' pre-miRNA homology (Kim et al. 2006).

Looking for miRNA targets for hsa-miR-1207-5p and hsa-miR-665, we found a total of 503 genes. From the total of targets, four were muscle specific AK1, CAMK2D and SLC25A34 which are regulated by hsa-miR-1207-5p and DIDO1, regulated by hsa-miR-665. The AK1 encodes for Adenylate kinase gene, which is an enzyme involved in regulating the adenine nucleotide composition within a cell by catalyzing the reversible transfer of phosphate group among adenine nucleotides. Three isoenzymes of Adenylate kinase have been identified in vertebrates, Adenylate isoenzyme 1 (AK1), 2 (AK2) and 3 (AK3). AK1 is found in the cytosol of skeletal muscle, brain and erythrocytes, whereas AK2 and AK3 are found in the mitochondria of other tissues including liver and heart. A significant increase was detected in the abundance and activity of AK1 in muscle from obese/overweight and morbidly obese women (Hittel et al. 2005). Here, this gene was identified as a target for the miRNA which had similarity with naturalized Brazilian breed Piau EST sequence, being this breed known for its ability to produce a large amount of fat.

For the same miRNA we have two more target genes (CAMK2D and SLC25A34). The CAMK2D, which is a Calcium/calmodulin-dependent protein kinase type II delta chain, is a multifunctional protein kinase, with complex structural and auto regulatory properties. Evidence suggests structural diversity in CaMKII isoform variants is an important determinant of cellular function (Hudmon et al., 2002 and Bayer et al., 2001). The SLC25A34 is a member of solute carrier family 25 which consists of proteins with function as transporters of a large variety of molecules (Pebay-Peyroula, et al., 2003). All these molecules originate from macromolecules that constitute the energy sources of cells and which are broken down into less complex molecules by cellular enzymes (Palmieri, 2004). The fourth gene, DIDO1, is a target of hsa-miR-665, and encode for Death inducer-obliterator 1 protein which is identified as a gene up regulated early in apoptosis by several stimuli. The over expression of DIDO-1 in cells induce massive apoptosis without any apoptotic stimuli (Garcia-Domingo et al., 1999).

Among these four genes identified, three and nine others which interact with them were added and could be grouped into a network of functional relevance which are: ADK, PKLR, PKM2, STAT1, CAMK2B, CREB1, GRIN2B, GRIA1, ACTN2, CAMK2D, AK1 and DIDO1. All proteins are connected with each other by the subnets kinase activity, synaptic transmission, cell projection, programmed cell death and genes involved in activation of NMDA receptor upon glutamate biding and postsynaptic events. The NMDA or N-methyl-D-aspartate, receptors are glutamate-gated ion channels widely expressed in the central nervous system which play key roles in excitatory synaptic transmission (Dingledine et al., 1999). This network can be helpful to better understand the genes molecular mechanisms.

The sequences which did not have similarity with miRNAs were aligned against others ncRNA. Most of the known ncRNAs fulfilled relatively generic functions in cells, such as the Ribosomal RNA (rRNAs) and transfer RNA (tRNAs) involved in mRNA translation, spliceosomal uRNAs or small nuclear RNAs (snRNAs) involved in splicing and small nucleolar RNAs (snoRNAs) involved in the modification of rRNAs (Mattick and Makunin, 2006). Some of the ncRNA classes identified here were the piRNA with the most expressive number of matches (38.3%); it is a class of small RNAs related to the mammalian germ line cells (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006). piRNAs interact with the Piwi subfamily of proteins and form a ribonucleoprotein complex called Piwi-interacting RNA complex (piRC), which has been extracted and purified from rat testes (Lau et al., 2006). Piwi gene was first discovered in reproductive stem cell in Drosophila, and it is an important regulating factor in stem cell division. Hiwi gene is a congener Piwi in human, and was first cloned from human testis cDNA library. Analysis of human adult samples showed that Hiwi was also expressed in a wide range of tissues including prostate, ovary, small intestine, brain, liver, skeletal muscle, heart, kidney and pancreas (Sharma et al. 2001). This may explain the presence of this RNA class in our data even though it has been more reported in germ line cells, but still their high level was intriguing (38.3%) being necessary more studies on it.

Another class identified with 1.2% of total matches was the small nucleolar RNAs (snoRNA) which is located within the nucleus. snoRNAs exist in the cell in the form of RNP complexes (snoRNPs), and have been discovered in protozoan, fungal, plant and mammalian cells (reviewed in Gerbi, 1995; Maxwell and Fournier, 1995; Smith and Steitz, 1997; Tollervey and Kiss, 1997). With the same number of matches than snoRNAs are the snRNAs (1.2%), they comprise a small group of highly abundant, non-polyadenylated, non-coding transcripts that function in the nucleoplasm (Matera et al., 2007). These classes exist as ribonucleoprotein complexes termed snRNPs (Guthrie and Patterson, 1988). Another class of RNA which had similarity with our sequences were the small non-messenger RNAs (snmRNAs). These molecules have cellular function on their own or as the snRNAs, in complex with proteins that are bounded to the RNA and thus form ribonucleoprotein complexes (Hüttenhofer et al. 2001).

Two important classes of RNA involved in mRNA translation are tRNAs and rRNAs, representing here 4.5 and 7.2% of the total matches, respectively. The first is responsible for the transport of a specific amino acid to a growing polypeptide chain during translation. The second is the RNA component of ribosome, they provide a mechanism to bring the necessary amino acids corresponding to the appropriate mRNA codon (Lodish et al. 2004). We also identified matches with non-coding sequences as introns (4.9%), 3'elements (2.9%) and internal ribosome entry sites (1.4%). With less than one percent we had sequences as gap stem loop 3, leader, protein biding sites, Rev response element, channel RNA editing signal, 3'UTR stem loop, P RNA, Psi packing element and Trans-activating regions (TAR). A large amount of matches (36.9%) were unclassified transcripts in the used data base. This may be explained by the high number of small RNAs which have not been posted in data bases yet (Jacquier, 2009).

For a better understanding of miRNAs biological role, we performed a qRT-PCR across seven periods (four prenatal and three postnatal) of three known miRNAs and two novels. The miRNAs miR-424 and miR-133a were described by Chen et al., 2008 when they had their expression levels measured during two muscle development periods (65 and 95 days of pregnancy). Their results showed that miR-424 is down-regulated while miR-133a is up-regulated across these two periods. Our results did not find this difference of expression level during this time, but a higher expression during prenatal stage was showed ($p<0.05$) for the gene miR-424, while the miR-133a was low expressed at both first periods getting higher from the 70th day of pregnancy until 171 days postnatal without varying statistically during this time (see **Table 6**). These genes are related to muscle development; so their high expression during prenatal stage suggests that they play an important role on muscle genes expression during this time. Other miRNAs analysed here were discovered in pigs by Zhou et al., 2010 (hsa-miR-147). They collected pig mRNAs from the NCBI GenBank nucleotide database and searched potential miRNAs based on similarity. The expression level in the skeletal muscle of this miRNA was analysed here during the mentioned periods showing a differential between pre- and postnatal stage ($p<0.05$) being more expressed during most of the prenatal stage, having high level again at 171 days post natal. This expression in the skeletal muscle has not been reported in the consulted literature.

In this work two potential new pig miRNAs were identified which had similarity with hsa-miR-1207-5p and hsa-miR-665. Only the hsa-miR-665 could be amplified and validated in pigs; however, these results are not definite once we used only commercial animals for expression analyses; the hsa-miR-1207-5p can be expressed in other pigs population such as the naturalized Brazilian breed Piau. The hsa-miR-665 expression levels during the mentioned periods showed a high expression during prenatal stage, when a peak during 70 and 90 days take place (see contrasts 40d/70d and 90/107 on **Table 6**). This period is known to be related with the second generation of muscle fiber formation which is between 54 to 90 days of pregnancy (Wigmore and Stickland, 1983). At 107 days postnatal, this expression decreases, making one see that this gene is highly expressed after the first generation of muscle formation (35-54 days) until birth, when it starts to be low.

Conclusion

This work suggests that EST analysis is a feasible strategy for identifying ncRNAs as novel miRNA candidates and their targets. In the current study, it was proposed the identification of two new pig miRNAs and other known non-coding RNAs through the homology search pipeline. The findings of miRNAs and their muscle-specific targets with their metabolic network will be helpful for understanding the function and processing of pigs small RNAs in the future. Furthermore, the expression level of the miRNAs analysed across pre- and postnatal stages demonstrate a role on biological events during skeletal muscle development and may be worthy of further investigation on the biological role in pigs.

Authors' contributions

SEFG and PSL conceived of the study, and participated in its design and coordination. CSN and LLV performed of the biological analysis. LLV and CSN were responsible for the bioinformatics and qRT-PCR analysis and wrote the manuscript with the help of SEFG. All authors have read and approved the final manuscript.

Acknowledgements

We would like to thank the Brazilian agencies FAPEMIG, CAPES, and CNPq for their financial support.

References

Ambros V. **The functions of animal microRNAs**. *Nature* 2004, 431: 350–355.

Ambros V., Bartel B., Bartel D.P. et al. **A uniform system for microRNA annotation**. *RNA* 2003, 9, 277–9.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T. **A novel class of small RNAs bind to MILI protein in mouse testes.** *Nature,* 2006, 442(7099):203-7

Bartel DP. **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**, 281−297.

Bayer K. U. and Schulman H. **Regulation of Signal Transduction by Protein Targeting: The Case for CaMKII**. *Biochemical and Biophysical Research Communication.* 2001, *289, 917–923.*

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, *et al*. **Identification of hundreds of conserved and nonconserved human microRNAs.** *Nature Genetics* 2005, **37**(7)**:**766-770.

Carrington JC, Ambros V. **Role of microRNAs in plant and animal development**. *Science* 2003, **301**(5631): 336–338

Carthew RW. **Gene regulation by microRNAs.** *Current Opinion in Genetics & Development* 2006, **16,** 203–208.

Chen J-H., Wei W-J., Xiao X., Zhu M-J., Fan B. and Zhao, S-H. **Expression Analysis of miRNAs in Porcine Fetal Skeletal Muscle on Days 65 and 90 of Gestation**. *The Asian-Australasian Journal of Animal Science* 2008, 21(7):954-960.

Garcia-Domingo D, Leonardo E, Grandien A, Martinez P, Albar JP, Izpisua-Belmonte JC, *et al*. **DIO-1 is a gene involved in onset of apoptosis *in vitro*, whose misexpression disrupts limb development**. *Proceedings of the National Academy of Sciences of the United States of America* 1999; 96: 7992–7.

Gerbi S.A. ***Small nucleolar RNA.*** Biochemistry and Cell Biology *1995,* ***73****, 845-858.*

Girard A, Sachidanandam R, Hannon GJ, and Carmell MA. **A germline-specific class of small RNAs binds mammalian Piwi proteins.** *Nature***,** 2006**, 442,** 199-202.

Griffiths-Jones S. **The microRNA Registry**. NAR 2004 32(Database Issue):D109-D111.

Grivna ST, Beyret E, Wang Z, Lin H. **A novel class of small RNAs in mouse spermatogenic cells.** *Genes and Development* 2006; **20**(13):1709-14.

Guerra-Assunção J.A. and Enright A.J. **MapMi: Automated Mapping of miRNA loci**. *BMC Bioinformatics* 2010, 11:133

Guthrie C and Patterson B. **Spliceosomal snRNAs.** *Annual Reviews of Genetics* 1988, **22**:387-419.

Huang, X. and Madan, A. **CAP3: A DNA sequence assembly program**. *Genome Research* 1999, **9**, 868-877.

Huang T-H, Zhu M-J, Li X-Y, Zhao S-H. **Discovery of Porcine microRNAs and Profiling from Skeletal Muscle Tissues during Development**. *PLoSONE* 2008, 3(9): e3225. doi:10.1371/journal.pone.0003225

Hudmon A. and Schulman H. **Structure–function of the multifunctional Ca^{2+}/calmodulin-dependent protein kinase II**. *Biochemical Journal* 2002, *364, 593–611*

Hüttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie J-P, Brosius J. **RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse.** *EMBO Journal* 2001, **20**, 2943 – 2953.

Jacquier, A. **The complex eukaryotic transcriptome: unex- pected pervasive transcription and novel small RNAs.** *Nature Reviews Genetics* 2009, **10**, (12):833–844.

Kim HJ, Cui XS, Kim EJ, Kim WJ, Kim NH. **New porcine microRNAs genes found by homology search**. *Genome* 2006, **49**: 1283–1286.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE**. Characterization of the piRNA Complex from Rat Testes**. *Science* 2006, **313** (5785): 363-367.

Lewis BP, Burge CB, Bartel DP. **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets**. *Cell* 2005, **120,** 15–20.

Li M, Xia Y, Gu Y, Zhang K, Lang Q, et al. **MicroRNAome of Porcine Pre- and Postnatal Development**. *PLoSONE* 2010, 5(7): e11541.doi:10.1371/journal.pone.0011541

Livak, K.J and Schmittgen, T.D. **Analysis of relative gene expression data using real-time quantitative PCR and the 2$^{-\Delta\Delta CT}$ method**. *Methods* 2001, **25**:402-408.

Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. **Molecular Biology of the Cell.** WH Freeman: New York, NY. 5th ed. 2004.

Markham, N. R. & Zuker, M. **DINAMelt web server for nucleic acid melting prediction**. *Nucleic Acids Research* 2005, **33**, W577-W581.

Markham, N. R. & Zuker, M. **UNAFold: software for nucleic acid folding and hybriziation**. In Keith, J. M., editor, *Bioinformatics, Volume II. Structure, Function and Applications* 2008, number 453 in *Methods in Molecular Biology*, chapter 1, pages 3–31. Humana Press, Totowa, NJ. ISBN 978-1-60327-428-9.

Matera AG, Terns RM, Terns MP. **Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.** *Nature Reviews Molecular Cell Biology* 2006, **8**, 209-220

Maxwell,E.S. and Fournier,M.J. **The small nucleolar RNAs**. *Annual Review of Biochemistry* 1995, **35**, 897–934.

Palmieri, F. **The mitochondrial transporter family (SLC25): physiological and pathological implications.** *Pflugers Arch.* 2004, **447**, 689–709.

Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trézéguet V, Lauquin GJ, Brandolin G. **Structure of mitochondrial ADP/ATP carrier in complex with Carboxyatractyloside**, *Nature* 2003, **426**, 39–44.

Pfaffl, M.W. **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic Acid Research* 2001, **29**, 900.

Rajewsky N. **microRNA target predictions in animals.** *Nature Genetics* 2006, **38** (Suppl), S8–13.

Seeman, S. E.; Gilchrist, M. J.; Hofacker, I. L.; Stadler, P. F. and Gorodkin, J. **Detection of RNA structures in porcine EST data and related mammals**. *BMC Genomics* 2007, **8:**316

Smith,C.M. and Steitz,J.A. **Sno storm in the nucleolus: new roles for myriad small RNPs**. *Cell* 1997, **89**, 669–672.

Tollervey,D. and Kiss,T. **Function and synthesis of small nucleolar RNAs**. *Curr. Opin. Cell Biol.*1997, **9**, 337–342.

Xie F L, Huang S Q, Guo K, et al. **Computational identification of novel microRNAs and targets in *Brassica napus***. *FEBS Lett* 2007, **581**, 1464-14739.

Zhang B H, Pan X, Cobb G P, et al. **Plant microRNA: A smallregulatory molecule with big impact**. *Development Biology* 2006, 289, 3-6.

Zhou, B. and Liu, H-L. **Computational identification of new porcine MicroRNAs and their targets**. *Animal Science Journal* 2010, **81,** 290-296.

Supplementary data

Table S1. EST's homologs mature miRNA location into a pig genome using Blast searches and MapMi analysis.

EST source	Mature miRNA sequence	Known miRNA	Chr.	Strand	Start position	End Position
Lw_contig39	GGGGGAGGGCCGGUGGCUGUCC	hsa-miR-149*	17	+	10565850	10565872
Lw_Contig135	CAUGGCUAUGUUUUCCUA	hsa-miR-4295	15	+	69325919	69326038
Lw_Contig83	UGAAUCCUUUCAGAGGAA	hsa-miR-516b*/hsa-miR-516a-3p	12	-	21678549	21678665
Lw_Uf_67B01	ACUCAAUCAUGGUGAAA	hsa-miR-4264	3	+	116103360	116103415
Pi_Contig104	UUGUGGGAGUGGCUGUU	hsa-miR-1275	13	-	116719052	116719176
Pi_Uf_30F11	UCAGAAAAUGUUUAUUAGAAAA	hsa-miR-545*	12	-	4233454	4233572
Pi_Uf_69F04	AUGUUAGAGGAAAAUACACUU	hsa-miR-376a/mmu-miR-376b/mmu-miR-376a	1	+	40950039	40950140
Pi_Uf_84A01[a]	UGGCCAGGAGGAUGGGAGGGU	hsa-miR-1207-5p	4	-	93723497	93723599
Lw_Uf_70D12	UUUCUUACCCUUCAUGAAGACU	hsa-miR-3685	1	+	246829732	246829754
Pi_Uf_76E10	UGAAAUGGAGAAGAGAUUUAAU	hsa-miR-3202	18	+	31286646	31286764
Pi_Uf_74A08	ACCAGUGAACUCUUCUCC	hsa-miR-1825	1	+	137005247	137005265
Pi_Uf_64E10	UGUGUGUGUGCGUGCGUGUA	mmu-miR-466i-5p	14	-	31065706	31065811
Lw_Uf_88G07/ Lw_Uf_80F11[b]	ACCAGCAGGCUGAAGCCCGG	hsa-miR-665	17	-	50090868	50090995
Lw_Uf_96H03	AGGAGGAGCUGGCUGAGUGCU	mmu-miR-1947	17	-	50090362	50090484

Chr., chromosome; EST Source: cDNA sequences which the miRNA was identified, Pi and Lw means that these sequences are from Brazilian naturalized Piau and Large White breed libraries respectively. **a** and **b** are the EST sequences which it were identified those novel miRNAs.

Table S2. miRNAs/endogenous primer concentration and cDNA. Their concentrations were determinate through amplification efficiency as described by Pfaffl, 2001.

Genes	Primer Concentration (nM)	cDNA concentration (ng)
miR-424	100	25
miR-133a	400	225
hsa-miR-1291	100	225
hsa-miR-147	100	75
hsa-miR-665	100	225
HPRT1	400	25

Chapter 4

GENERAL DISCUSSION

In the previous chapter we have seen the importance of knowledge of muscle development for meat production as the presence of regulatory elements playing important roles on many biological processes. The use of expressed sequence tags (EST) provides direct information about the transcriptome and indirectly, on the relationship between the genome and different phenotypes, supplying knowledge about differentially expressed genes (DEG) and sequences transcribed for the control of gene expression, e.g., some non-coding RNAs. Characteristics of muscle tissue in pigs may have been directly influenced by both, genes and non-coding RNAs, in different stages of development.

In the second chapter the comparison of three sequenced cDNA libraries from divergent genetic background identified sequences that were differentially expressed among them. From 34 differentially distributed gene sequences identified, 21 could be grouped into a network of functional relevance for muscularity; this may be helpful to better understand the molecular mechanisms involved in skeletal muscle development. However, this is a complex trait that is subject to action from a large number of genes that are regulated by several transcription regulatory elements (Carvajal & Rigby, 2010). To better understand the DEGs role during muscle development, their expression levels were measured during different muscle periods. Thus it was possible to confirm genes which are known to be related with myogenic stage as ANKRD2, MYBPC1, NEB and MYL2 (Bean et al. 2008; Gautel et al., 1998 & Kurasawa et al., 1999; Bang et al. 2006 and Zhang et al. 2009 respectively) having a prenatal high expression in this study. Furthermore, novel candidates for muscle development (TPM2, TP53 and DCTN1) can be listed.

In the third chapter the use of bioinformatics tools proved to be very useful. We made use of the property of miRNAs to be highly conserved between closely related species in order to predict novel pig miRNA candidates (Kim et al., 2006). A Blast pipeline suggested the discovery of two novel miRNA by computational approaches following the criteria proposed by Ambros et al., 2003. A substantial number of remaining no hit sequences showed homology with non-coding sequences as Piwi-interacting RNAs (piRNA) and others, but still there is a large amount of sequences without homology that may be explained by the high number of small RNAs which were not posted in data bases yet (Jacquier, 2009).

The two miRNA identified showed their mature sequence similar to hsa-miR-1207-5p and hsa-miR-665 but their precursor sequences were not, when compared with their supposed miRNA families. That is, the precursor sequences of the pig miRNAs identified here are poorly conserved according to studies that showed the occurrence of non conserved miRNAs and their precursors in mammals and plants (Bentwich et al. 2005 and Li et al. 2010). Looking for miRNA targets we found multiples genes. From the total of targets, four were muscle specific (AK1, CAMK2D, SLC25A34 and DIDO1) and are involved in biological process like nucleobase, nucleoside, nucleotide and nucleic acid metabolic; regulation of cell growth; transport and programmed cell death, respectively.

The expression analyses were performed for three known miRNAs (miR-424, miR-133a and hsa-miR-147) and two novel in pigs (hsa-miR-1207-5p and hsa-miR-665) across seven periods. From those two potential novel pig miRNAs, only hsa-miR-665 could be amplified and validated in pigs. However, these results are not definite once we used only commercial animals for expression analyses; the ssc-miR-new1 can be expressed in other pigs population such as naturalized Brazilian breed Piau. The hsa-miR-665 expression levels showed a high expression during prenatal stage covering important periods of muscle development.

These studies may provide valuable information allowing a better understanding of the molecular mechanisms involved in muscle development. Analyses of DEG in the pre-and postnatal life suggest the presence of genes acting specifically on certain stages of muscle development by contributing to better explain their functions. Furthermore, the identification of two new miRNAs, together with other previously identified and posted on the databases in pigs, may contribute to a better understanding regarding the gene regulation mechanisms and its importance for genetics studies and animal breeding, allowing the understanding of the physiology deposition of muscle to meat production in pigs.

REFERENCES

Ambros V, Bartel B, Bartel DP et al. (2003). **A uniform system for microRNA annotation**. *RNA*, 9, 277–9.

Bang ML, Li X, Littlefield R, Bremner S et al. (2006). **Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle.** *The Journal of Cell Biology* 2006, **173**:905-916.

Bean C, Facchinello N, Faulkner G, Lanfranchi G: **The effects of Ankrd2 alteration indicate its involvement in cell cycle regulation during muscle differentiation.** *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2008, **1783**(6):1023-1035.

Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, *et al*. **Identification of hundreds of conserved and nonconserved human microRNAs.** *Nature Genetics 2005*, 37(7)**:**766-770.

Gautel M, Furst DO, Cocco A, Schiaffino S: **Isoform transitions of the myosin binding protein C family in developing human and mouse muscles: lack of isoform transcomplementation in cardiac muscle.** *Circulation Research* 1998, **82**:124-129.

Jacquier A. **The complex eukaryotic transcriptome: unex- pected pervasive transcription and novel small RNAs.**_*Nature Reviews Genetics 2009*, 10, (12):833–844.

Kim HJ, Cui XS, Kim EJ, Kim WJ et al. (2006). **New porcine microRNAs genes found by homology search**. *Genome*, 49:1283–1286.

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

Lewis BP, Burge CB, Bartel DP. **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets**. *Cell* 2005, 120:15–20.

Li M, Xia Y, Gu Y, Zhang K, Lang Q, et al. **MicroRNAome of Porcine Pre- and Postnatal Development**. *PLoSONE* 2010, 5(7): e11541.doi:10.1371/journal.pone .0011541

Lu YZ and Liu J. ***In silico* identification of MicroRNAs and their targets in diatoms.** *African Journal of Microbiology Research* 2010, 4(13):1433-1439

Rajewsky N. **microRNA target predictions in animals.** *Nature Genetics* 2006, 38 (Suppl), S8–13.

Zhang S-Z, Xu Y, Xie HQ, Li X-Q, Wei W-Q, Yang Z-M: **The possible role of Myosin light chain in myoblast proliferation.** *Biology Research*. 2009, **42**:121-132.