BRUNA PENA SOLLERO

TRANSCRIPTIONAL PROFILING DURING PIG SKELETAL MUSCLE DEVELOPMENT OF A BRAZILIAN LOCAL BREED AND TWO COMMERCIAL GENETIC GROUPS

Thesis presented to the Animal Science Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of *Doctor Scientiae.*

VIÇOSA
MINAS GERAIS - BRASIL
2010

BRUNA PENA SOLLERO

TRANSCRIPTIONAL PROFILING DURING PIG SKELETAL MUSCLE DEVELOPMENT OF A BRAZILIAN LOCAL BREED AND TWO COMMERCIAL GENETIC GROUPS

Thesis presented to the Animal Science Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of *Doctor Scientiae*.

APROVADA: October 19, 2010.

Prof. Fabyano Fonseca e Silva
(Co-adviser)

Prof. Flávia Maria Lopes Passos

Prof. Paulo Sávio Lopes
(Co-adviser)

Prof. Catherine W. Ernst

Prof. Simone Eliza Facioni Guimarães
(Adviser)

"Do not go where the path may lead.
Go instead where there is no path
and leave a trail"

- Ralph Waldo Emerson

ACKNOWLEDGEMENTS

To *Universidade Federal de Viçosa* and the Animal Science Department, for providing me the opportunity to conclude an important step in my life;

To *CNPq*, for the scholarship here and in United States (fellowship program);

To my parents and sisters, for all their affection and unconditional support;

To my adviser, Professor Simone Eliza Facioni Guimarães, for her support, for enriching my knowledge, increasing my confidence and being my confirmation that "a good example is the best sermon";

To my co-adviser Fabyano Fonseca, for all the time he spent giving me fundamental advice that enabled me to develop and complete this work;

To Professor Paulo Sávio for his guidance while co-advising me;

To Professor José Domingos Guimarães, for his essential support;

Many thanks to the Department of Animal Science at Michigan State University, the USDA CSREES National Research Initiative, and Michigan Agricultural Experiment Station;

To Dr. Ernst for the opportunity to work with her research group and develop part of this project at the Meat Animal Muscle Biology and Molecular Genetics Laboratory (MSU);

To Dr. Steibel and Dr. Tempelman;

To Valencia, Nancy, IgSeo, Michelle and all my friends in US (Julia, Kristi, Oscar, Julie, Rita, Erik, Koushik) for their company, help, friendship and patience;

To Professor Glória Regina Franco and graduate student Priscila Grinberg at the *Departamento de Genética Bioquímica* of the *Universidade Federal de Minas Gerais*, for the opportunity to create the third step of this project and for their collaboration;

To all the others Professors, employees and students from the Animal Science Department, for making workdays a pleasant living environment, sharing ideas and enjoying great moments;

To my friends from LABTEC, for all these years of friendship and complicity: Débora, Carlos, Lincoln, Kleibe, Katiene, Ana Paula, Renata, Priscila, Lucas, Erica, Mayara and Carol; also Nicola and Miguel. Specially, many thanks to dear Marcos, for his willingness to participate- his collaboration was essential!

To all my other friends and relatives, for sharing with me happiness, fears, victories and overruns!

And finally, for finding God's presence in every day of my life!

BIOGRAPHY

Bruna Pena Sollero, daughter of Paulo Antônio Sollero and Dolores Maria Pena Sollero, born on September 1, 1981 in Brasilia, *Distrito Federal*, Brazil. She began her studies in Animal Science at the *Universidade Federal de Viçosa* in 2000, studying many different areas and participating in many different research groups and projects, especially with the Program for Breeding Swine.

In January of 2005, she chose to major in Animal Science, began pursuing a Master of Agricultural Science at the *Universidade de Brasília*, Brazil, in March of that same year. This degree was completed in partnership with *Embrapa Recursos Genéticos e Biotecnologia* (Embrapa Genetic Resources and Biotechnology) in the area of Animal Genetic Resources Conservation. Ms. Sollero completed her final examination on November 10, 2006 to obtain the title of Magister Scientiae in Animal Production.

In March of 2007 she began the doctoral graduation program in Animal Breeding at the Animal Science Department of the *Universidade Federal de Viçosa*. During her doctoral program she spent one year, from September of 2008 to September of 2009, at the Meat Animal Muscle Biology and Molecular Genetics Laboratory at Michigan State University in the United States. It was there that she developed a large part of her doctoral research project. Ms. Sollero completed her dissertation defense on October 19, 2010, to obtain the title of Doctor Scientiae in Animal Science.

SUMMARY

ABSTRACT

SOLLERO, Bruna Pena, D.Sc., Universidade Federal de Viçosa, October, 2010. **Transcriptional profiling during pig skeletal muscle development of a Brazilian local breed and two commercial genetic groups**. Adviser: Simone Eliza Facioni Guimarães. Co-Advisers: Paulo Sávio Lopes e Fabyano Fonseca e Silva.

Skeletal muscle development is a complex process involving the coordinated expression of thousands of genes. The aim of this study was to identify differentially expressed genes in *longissimus dorsi* (LD) muscle of pigs and their expression pattern through five time-points of the development. Firstly, a microarray analysis using the Pigoligoarray containing 20,400 oligonucleotides was done to compare two genetic groups of pigs that differ in muscularity (North American Yorkshire x Landrace (YL) crossbred pigs and Piau pigs -a naturalized Brazilian breed) at 40 and 70 days (d) of gestation (developmental stages encompassing primary and secondary fiber formation). A total of 486 oligonucleotides were differentially expressed (FC $\geq$ 1.5; FDR $\leq$ 0.05) between 40 and 70 d gestation in either YL or Piau pigs, and a total of 1,300 oligonucleotides were differentially expressed (FC $\geq$ 1.5; FDR $\leq$ 0.05) between YL and Piau pigs at either age. Gene ontology annotation and pathway analyses determined functional classifications for differentially expressed genes and revealed breed-specific developmental expression patterns. Thirteen genes were selected for confirmation by qRT-PCR analyses and expression patterns for most of these genes were confirmed, providing further insight into the roles of these genes in pig muscle development. The relative abundance of transcripts tended to be greater for the Piau pigs at 70 d of gestation suggesting that gene expression in LD muscle of YL pigs may be more delayed than in Piau pigs. The second experiment focused to evaluate the expression profile during prenatal (21, 40, 70 and 90 days pos conception) and early postnatal (10 weeks) stages of the same muscle between the local Brazilian pig breed (Piau) and a Brazilian white composite line. Based on qRT-PCR analyses, others fourteen genes related to intrinsic biological processes during myogenesis were investigated, and nine genes out of them were differentially expressed between both genetic groups ($P<0.05$). Significant different levels of expression were also observed through the five time points and patterns of genes related to cell proliferation, cell differentiation, energy metabolism, and maintenance were distinct in a breed-specific way. A new reference gene for expression profile investigation of pig muscle not used before for this purpose is proposed: DDIT3 (DNA-damage-inducible transcript 3). Furthermore, this research evaluated the application of three different protocols for data normalization and assessment of genes differentially expressed in microarray analyses. A trade-off among the results was observed when comparing the three protocol analyses tested by distinct statistical methods. The two new protocols proposed, in comparison with the first

microarray analysis applied in the first chapter, detected genes as DE only in two (C40vsP40 and P40vsP70) out of the four contrasts under evaluation (C40vsP40, P40vsP70, C70vsP70 and C40vsC70). The results strengthened the idea that the Robust Spline normalization method was able to identify more genes differentially expressed (FDR-False Discovery Rate ≤0.05) related with important GO terms involved with the muscle development process (especially in breed-contrasts) in comparison with the Loess method. In addition, normexp background correction method was advisable, working clearly better than the traditional subtraction method. It was suggested that the *q-value* FDR method can be too flexible to detect genes differentially expressed in small microarray experiments, while the BH FDR method can be too restrictive for that – although pointing out genes DE with higher fold change values and with similar functionalities related to the myogenesis process in context. In general, this study which was the first whole-genome expression evaluation of the Brazilian native Piau pigs, revealed both developmental and breed-specific patterns of gene expression in fetal pig skeletal muscle including genes not previously associated with myogenesis, and these information can contribute to future pig genetic improvement efforts. Also, comparisons among protocol analyses for microarray data are suggested for cleaning and improving the quality of the measures of gene expression.

RESUMO

SOLLERO, Bruna Pena, D.Sc., Universidade Federal de Viçosa, outubro de 2010. **Perfil transcricional durante o desenvolvimento muscular esquelético de suínos de uma raça local Brasileira e dois grupos genéticos comerciais.** Orientadora: Simone Eliza Facioni Guimarães. Co-orientadores: Paulo Sávio Lopes e Fabyano Fonseca e Silva.

O desenvolvimento esquelético muscular é um processo complexo que envolve a expressão coordenada de milhares de genes. O objetivo deste estudo foi identificar genes diferencialmente expressos no músculo *longissimus dorsi* (LD) de suínos, bem como seus padrões de expressão entre cinco estádios do desenvolvimento. Primeiramente, uma análise de microarranjo utilizando o *Pigoligoarray*, contendo 20.400 oligonucleotídeos, foi realizada para comparar dois grupos genéticos de suínos que diferem em capacidade de crescimento muscular e conteúdo de carne magra (grupo cruzado comercial Yorkshire-Landrace-YL, proveniente dos Estados Unidos e suínos Piau- raça Brasileira naturalizada) aos 40 e 70 dias (d) de gestação (estádios do desenvolvimento decorrentes da formação de fibras musculares primárias e secundárias). Um total de 486 oligonucleotídeos foram diferencialmente expressos ($FC \geq 1.5$; $FDR \leq 0.05$) entre 40 e 70 dias de gestação no grupo YL ou Piau, e um total de 1.300 oligonucleotíedos foram diferencialmente expressos ($FC \geq 1.5$; $FDR \leq 0.05$) entre suínos YL e Piau em ambas as idades pré-natais. Anotações baseadas no *Gene Ontology* (GO) e análises de vias metabólicas determinaram classificações funcionais para os genes diferencialmente expressos e revelaram padrões de expressão raça-específicos. Treze genes foram selecionados para confirmação por análises de qRT-PCR e os padrões de expressão para a maioria destes genes foram confirmados, gerando informações relevantes sobre seus papéis desempenhados durante o processo do desenvolvimento muscular de suínos. Na raça Piau a abundância relativa de transcritos aos 70 d de gestação foi tendencialmente maior, sugerindo que a expressão gênica no músculo LD de suínos YL pode ser atrasada em relação à raça local. O segundo experimento objetivou avaliar perfis de expressão durante cinco períodos pré-natais (21, 40, 70 e 90 dias pós-concepção) e um período pós-natal (10 semanas pós-nascimento) no mesmo músculo esquelético (LD) entre a raça Brasileira local Piau e uma linhagem comercial também brasileira. Baseado em análises de qRT-PCR, outros quatorze genes relacionados a processos biológicos intrínsecos ao processo de miogênese foram investigados, e nove destes genes foram significativamente ($P<0.05$) diferencialmente expressos entre ambos os grupos genéticos de suínos. Níveis significativos de expressão gênica também foram observados entre os cinco estádios avaliados e padrões de genes relacionados à proliferação celular, diferenciação celular, metabolismo de energia e manutenção da formação muscular foram distintos de forma raça-específica. Um novo gene referência, não utilizado como tal previamente em estudos de expressão muscular de suínos, foi

proposto: DDIT3 (DNA-damage-inducible transcript 3). A última etapa desta pesquisa objetivou avaliar a aplicação de três protocolos para normalização e detecção de genes diferencialmente expressos em análises de microarranjos. Resultados conflitantes foram observados quando comparado os três protocolos analisados por distintos métodos estatísticos. Os dois protocolos propostos, em relação à primeira análise de microarranjo aplicada no primeiro capítulo, detectaram genes como diferencialmente expressos em apenas dois (C40vsP40 e P40vsP70) dos quarto contrastes sob avaliação (C40vsP40, P40vsP70, C70vsP70 e C40vsC70). Os resultados demonstraram que o método de normalização *Robust Spline* foi capaz de identificar um maior número de genes diferencialmente expressos (FDR-Taxa de Falsas Descobertas ≤ 0.05) relacionados a importantes *GO terms* envolvidos no processo de desenvolvimento muscular (especialmente em contrastes entre idades) em comparação ao método Loess. Além do mais, o método de correção *background normexp* foi recomendado, uma vez que se mostrou mais acurado do que o método tradicional de Subtração. Foi sugerido que o método FDR *q-value* pode ser muito flexível na detecção de genes diferencialmente expressos em pequenos experimentos de microarranjo, enquanto o método BH pode ser muito restritivo – ainda que capaz de apontar genes DE com maiores valores de expressão relativa e funcionalmente relacionados ao processo de miogênese em contexto. Em geral, este estudo, que pela primeira vez avaliou em nível genômico a expressão de genes da raça Piau, revelou padrões de expressão gênica raça e idade-específicos durante o desenvolvimento muscular fetal, incluindo genes não descritos previamente como associados a tal processo. Tais informações podem contribuir para futuros avanços do melhoramento genético de suínos. Também, comparações entre diferentes protocolos de análises de microarranjos foram sugeridas, a fim de melhorar a qualidade e representar de forma mais clara as medidas de expressão gênica.

General Introduction

According to Abipecs (*Associação Brasileira da Indústria Produtora e Exportadora de Carne Suína*, 2010), the growing demand for protein food sources is a worldwide reality and the pork is continuously increasing its importance on the meat market.

For the meat-producing industries, emphasis has been given on increasing the understanding of the physiological process and molecular pathways associated with skeletal muscle growth and development (Reecy et al., 2006). In addition to the knowledge that factors operating prenatally theoretically determine the maximum size to which this tissue can growth (Wigmore & Stickland, 1983), skeletal muscle is essential for a wide range of functions in livestock animals, including breathing, locomotion, maintenance of posture and thermogenesis (Dauncey & Gilmour, 1996).

Pig breeding companies are focused on the organization and genetic structures of "artificial pig populations"; defining specific breeding goals, their weighting in an overall breeding objective and the choice of optimal selection criteria. Challenges imposed to these companies resulted from the evolution of genetic improvement combined with constant changing in the production systems. As discussed by van der Steen et al., (2005), in particular, highly heritable traits, such as milk production in dairy cows and lean percentage in pigs and broilers, are predominantly improved through the genetic route. Nevertheless, behind each choice of selection criteria strong researches focusing on the molecular genetics aspects of a certain population is constantly under evaluation.

The process of domestication, as described for instance by Jonsson (1991), includes selective breeding for specific characteristics, and may be considered as a first step in genetic improvement. An idea of the distance covered from this primary process can be gained by comparing the modern domestic pig to its wild counterpart. In this scenario, genetic groups, which were not submitted to the same selection process or still maintain the original genetic bases from the beginning, are attractive genetic reservoirs to investigate economically important characteristics in the swine species.

The Piau Brazilian naturalized pig breed, originated from other breeds introduced by Portuguese settlers in the 16th century, is also influenced by Dutch and African pig breeds (Vianna, 1985). These animals are considered to be fat-type, used to be reared in small farms, supplying farmers with meat and a large amount of fat. Piau was the first Brazilian breed to be registered at *Associação Brasileira de Criadores de Suínos* in 1986, and is one of the most recognized local breeds in Brazil (Mariante et al., 2003; Sollero et al., 2008). Since 1998 the Pig Breeding Farm at *Universidade Federal de Viçosa* allocates Piau pigs for research purposes and many trials have been carried out using this breed, such as QTL mapping using divergent crosses (Guimarães & Lopes, 2000; Pires et al., 2008; Paixão et al., 2008; Silva et al., 2009) and association studies between expression of candidate genes and intramuscular content (Serrão et al., 2009).

The pig genome has been analyzed to address biomedical, agricultural and fundamental biological questions, using more and more sensitive and comprehensive tools. The first draft of the domesticated pig genome announced in November of the last year increased even more the perspectives (http://www.thaindian.com/newsportal/health/first-draft-of-the-swine-genome-sequence-revealed_100269189.html#ixzz0zKVWsYZj). We know that there is a finite number of genes for pigs (approximately 30,000), a number though that is still very large, and require more detailed studies in the transcriptome of this specie. Great advances in our understanding of the porcine transcriptome have occurred over the past decade, and especially in the past few years (Tuggle et al., 2007). One of the most common tool used to carry out the measurement of RNA transcript in a particular tissue, for instance, include complementary oligonucleotide microarrays (Barrett & Kawasaki, 2003).

Although the increased number of publications, level of execution and interpretation of microarray experiments, researchers are still in the very early stages of understanding RNA expression profiles considering complex biological processes such as myogenesis. In general, there are many protocols and types of systems available, but the basic technique of microarray involves extraction of RNA from biological samples in either normal or interventional states (Butte, 2002).

In the case of two-color microarrays, the RNA (or, in some protocols, isolated messenger RNA) under evaluation is stained with different fluorescences in order to distinguish them as control and tested group. The labeled cDNA/cRNA groups are then hybridized to a microarray slide which contains probes designed to represent the transcriptional variety of the specific tissue or biological system under investigation (Hoen et al., 2003). When the hybridization step finish, the excess is washed off and the microarray is scanned under laser light detecting specific absorbance waves which designate each fluorescence. In oligonucleotide microarrays, all probes have been designed to be theoretically similar regarding to hybridization temperature and binding affinity. Each microarray measures each contrasting samples and provides an absolute measurement level for each RNA molecule hybridized to each probe. The end of the process results in thousand measurements of gene expression per biological samples simultaneously.

Theoretically, the oligonucleotides synthesized to compose microarray platforms have very similar melting temperatures or G-C (guanosine – cytosine) content, very little homology with other oligonucleotides, are entirely contained within an exon, and have no repetitive – or harping sequences and other characteristics (Stears et al., 2003; Woo et al., 2004; Zhao et al., 2005; Sethi et al., 2008) which justify the adoption of oligonucleotide microarrays.

In accordance with White & Salamonsen, (2005) the main disadvantage of using oligonucleotide microarrays is their high cost. However, initiatives like The Swine Protein-Annotated Oligonucleoide Microarray, has been developed as open source collaboration between investigators and institutions with an interest in pig physiology. The sequences of the oligonucleotides and the annotation of the consensus sequences are provided at no cost to the entire research community. In this manner, this type of array become economically attractive, and thus, constantly used in research. Based on a novel 70-mer oligonucleotide microarray for profiling expression of the pig (*Sus scrofa*) genome, this microarray platform is enable to proceed rapid and simultaneous comparison of mRNA levels for thousands of genes (Steibel et al., 2009).

The main issue in microarray experiments is data analysis and the consequent extraction of biological knowledge. Transcriptome analysis itself is complicated by multiple factors such as the limited number of possible experiment replications - which is always lower than the number of variables, i.e. genes under investigation (Draghici et al., 2001) - and the limited knowledge of gene regulation and gene product function in different conditions or systems. Moreover, the multiplicity of problems in which thousands of hypotheses are tested simultaneously within one experiment (Dudoit et al., 2002) is another tricky characteristic in microarray analyses. Because of that, it is also advisable to compare different protocol analyses of microarray experiments with the objective of cleaning and improving the quality of the measures of gene expression.

Whatever information embedded in a microarray experiment appears to be entangled in a complex mix of various types of noise. This has caused some researchers to call for establishing industrial manufacturing standards through validation of the technology (Verducci et al., 2006). Others welcome the diversity of platforms and analytic methods as complementary forms of discovery, relying on alternative PCR-based technologies for validation of expression levels. Quantitative real time PCR analysis is being a feasible way to address and confirm the differential expression of genes with more sensitivity and specificity. The qRT-PCR is a complex assay which all physical and chemical components of the reaction are interdependent (Bustin, 2000). Based on the reverse transcription of the RNA template into cDNA, followed by its exponential amplification in a traditional PCR reaction, there are a wide range of choices of methodologies differing in processivity, fidelity, thermal stability and ability to read the template strand. The SYBR® Green is one of the methods of choice due to inexpensive (particular to each Lab) and sensitivity. It is a fluorogenic dye that exhibits little fluorescence when in solution, but emits a strong fluorescent signal upon binding to double-stranded DNA after a primer pair annealing. The Taqman® method (Perkin-Elmer-Applied Biosystem) utilizes the 5'-nuclease activity of the DNA polymerase to hydrolyse a hybridization probe bound to its target amplicon (together with the primer, after denaturation) during PCR, which reflects the high specificity of the technic.

Introducing the biological process to be considered in this research, basically, the development of skeletal muscle fibers during fetal growth (myogenesis) in mammals takes place in two waves known as primary and secondary fiber formation (Wigmore & Evans, 2002). During each wave, myoblasts proliferate and fuse to form new muscle fibers. Primary fibers are formed de novo, whereas secondary fibers form around a primary fiber. In pigs, this process takes place at approximately 30-60 days of gestation and 54-90 days of gestation for primary and secondary fibers, respectively (Wigmore & Stickland, 1983).

The goal of this study was to use whole-genome microarray and quantitative real time PCR molecular technologies to identify both developmental and breed-specific patterns of gene expression in the *longissimus* dorsi (LD) muscle of pig breeds (or genetic groups) with distinctly different muscularity. Skeletal muscle development involves the synchronized expression and interaction of many genes, and this study aims to provide additional insight into the process of muscle development, covering the early stages of myogenesis until the beginning of postnatal stage development of pigs. We also discussed the application of different protocol analyses in microarray data, in order to investigate specific steps of the analyses that may be decisive in the generation and interpretation of reliable results.

References

Barrett, JC and ES Kawasaki (2003) Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Drug Discov Today* **8**:134-41.

Bustin, SA (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**:23-39.

Butte, A (2002) The use and analysis of microarray data. *Nature Reviews* **1***:* 951-960.

Dauncey, MJ and Gilmour RS (1996) Regulatory factors in the control of muscle development. *Proceedings of the Nutrition Society* **55**: 543-559.

Draghici S., AK, B Ho and S Shams (2001) Experimental design, analysis of variance and slide quality assessment in gene expression arrays. *Current Opinion in Drug Discovery and Development*, **4** (3):332-337.

Dudoit, S, YH Yang, MJ Callow and TP Speed (2002) Statistical method for identifying with differential expression in replicated cDNA microarray experiments. *Statistica* Sinica, **12**: 111-139.

Guimarães, S. E. F., Lopes, P. S. (2000) Use of native genetic resources as a tool for the genomic mapping in swine. In: Rare Breeds International Symposium, Brasília, DF, Brazil.

Hoen, PAC, F de Kort, GJB van Ommen and JT den Dunnen (2003) Fluorescent labelling of cRNA for microarray applications. *Nucleic Acids Research* 31: e20. doi: 10.1093/nar/gng020.

Jonsson P (1991) Evolution and domestication, an introduction. In: K Maijala (ed) *Genetic Resources of Pig, Sheep and Goat*. *World Animal Science*, **12**, Elsevier, Amsterdam, pp. 1-10.

Mariante, A. S., Castro, S. T. R., Albuquerque, M. S. M., Paiva, S. R., Germano, J. L. (2003). Pig Biodiversity in Brazil. *Archivos de Zootecnia da Universidade de Córdoba* **52:** 245-248.

Paixão D.M., Guimarães S.E.F., Silva Filho M. I., Lopes P.S, Pereira M.S., Sollero B.P. (2008) Detecção de locos de características quantitativas nos cromossomos 16, 17 e 18 de suínos. Brazilian Journal of Animal Science 37, 1781-87.

Pires, A. V., Lopes, P. S., Guimarães, S. E. F., Guimarães, C. T., Peixoto, J. O. (2008). Mapeamento de locos de características quantitativas, no cromossomo seis suíno, associados às características de cortes de carcaça. Arquivo Brasileiro de Medicina Veterinária e Zootecnia, 60, 725-732.

Reecy, JM, DM Spurlock and CH Stahl (2006) Gene expression profiling: Insights into skeletal muscle growth and development. *Journal of Animal Science* **84**: 150-154.

Serão, NVL, (2009) Association of candidate gene expression with intramuscular fat content in the porcine longissimus dorsi muscle. M.Sc. dissertation, Universidade Federal de Viçosa: Viçosa.

Sethi, D, A Kummar, KC Gupta and P Kumar (2008) A facile Method for the Construction of Oligonucleotide Microarrays. *Bioconjugate Chem* **9**: 2136–2143.

Silva, K. M. , Guimarães, S. E. F. , Lopes, P. S. , Nascimento, C. S. , Lopes, M. S. , Weller, M. M. d C. A. (2009). Mapeamento de Locos de Características Quantitativas para Desempenho no Cromossomo 4 de suínos. Revista Brasileira de Zootecnia / *Brazilian Journal of Animal Science* **38**, 474-479.

Sollero B.P., Paiva S.R., Faria D.A., Guimarães S.E.F., Castro S.T.R., Egito A.A., Albuquerque M.S.M., Piovezan U. Bertani G.R. & Mariante A. da S. (2008) Genetic diversity of Brazilian pig breeds evidenced by microsatellite markers. *Livestock Science* **123**: 8-15.

Stears, RL, T Martinsky and M Schena (2003) Trends in microarray analysis. *Nature Medicine* **9**: 140-145.

Steibel J.P., Wysocki M., Lunney L.K., Ramos A.M., Hu Z.-L., Rothschild M.F. & Ernst C.W. (2009) Assessment of the swine protein-annotated oligonucleotide microarray. *Animal Genetics* **40**: 883-93.

Tuggle, CK, Y Wang and O Couture (2007) Advances in Swine Transcriptomics. *International Journal of Biological Science* **3**:132-152.

van der Steen, HAM, GFW Prall and GS Plastow (2005) Application of genomics to the pork industry. *Journal of Animal Science* **83**: 1-8.

Verducci, JS, VF Melfi, S Lin, Z Wang, S Roy and C K Sen (2006) Microarray analysis of gene expression: considerations in data mining and statistical treatment. *Physiology Genomics* **25**: 355–363.

Vianna, A. T. (1985). *Os suínos*. (14th ed.). São Paulo, Brasil: Nobel.

White, CA and LA Salamonsen (2005) A guide to issues in microarray analysis: application to endometrial biology. *Reproduction* **130:**:1–13. doi: 10.1530/rep.1.00685.

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

Wigmore P.M. & Evans D.J. (2002) Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis. *International Review of Cytology* **216:** 175-232.

Woo, Y, J Affourth, S Daigle, A Viale, K Johnson, J Naggert and G Churchill (2004) A comparison of cDNA, Oligonucleotides and Affymetrix GeneChip Expression Microarray Platforms. *Journal of Biomoleclar Techniques* **15**: 276-284.

Zhao, S-H, J Recknor, JK Lunney, D Nettleton, D Kuhar, S Orley, CK Tuggle (2005) Validation of a first-generation long-oligonucleotide microarray for transcriptional profiling in the pig. *Genomics* **86**: 618- 625.

CHAPTER I
(ARTICLE I)

Transcriptional profiling during fetal skeletal muscle development of Piau and Yorkshire-Landrace crossbred pigs

B.P. Sollero[*,†], S.E.F. Guimarães[†], V.D. Rilington[*], R.J. Tempelman[*], N.E. Raney[*], J.P. Steibel[*], J.D. Guimarães[‡], P.S. Lopes[†], M.S. Lopes[†] and C.W. Ernst[*]

[*]Department of Animal Science, Michigan State University, East Lansing, 48824, USA; [†]Animal Science Department, Universidade Federal de Viçosa, Viçosa, 36571-000, Brazil; [‡]Veterinary Department, Universidade Federal de Viçosa, Viçosa, 36571-000, Brazil

Abstract

Skeletal muscle development is a complex process involving the coordinated expression of thousands of genes. The aim of this study was to identify differentially expressed genes in *longissimus dorsi* (LD) muscle of pigs at 40 and 70 days (d) of gestation (developmental stages encompassing primary and secondary fiber formation) in Yorkshire-Landrace (YL) crossbred pigs and Piau pigs (a naturalized Brazilian breed), two breed types that differ in muscularity. Fetuses were obtained from gilts at each gestational age (n = 3 YL; n = 4 Piau), and transcriptional profiling was performed using the Pigoligoarray microarray containing 20,400 oligonucleotides. A total of 486 oligonucleotides were differentially expressed (FC ≥ 1.5; FDR ≤ 0.05) between 40 and 70 d gestation in either YL or Piau pigs, and a total of 1,300 oligonucleotides were differentially expressed (FC ≥ 1.5; FDR ≤ 0.05) between YL and Piau pigs at either age. Gene ontology annotation and pathway analyses determined functional classifications for differentially expressed genes and revealed breed-specific developmental expression patterns. Thirteen genes were selected for confirmation by qRT-PCR analyses and expression patterns for most of these genes were confirmed, providing further insight into the roles of these genes in pig muscle development. This study revealed both developmental and breed-specific patterns of gene expression in fetal pig skeletal muscle including genes not previously associated with myogenesis, and this information can contribute to future pig genetic improvement efforts.

Keywords: microarray, myogenesis, pig, qRT-PCR

Resumo

O desenvolvimento do músculo esquelético é um processo complexo que envolve a expressão coordenada de milhares de genes. O objetivo deste estudo foi identificar genes diferencialmente expressos no músculo *Longissimus dorsi* de suínos aos 40 e 70 dias de gestação (estádios do desenvolvimento decorrente da formação de fibras primárias e secundárias) em uma raça comercial resultante do cruzamento entre Yorkshire-Landrace (YL) e suínos Piau (raça naturalizada Brasileira), dois grupos genéticos divergentes em musculosidade. Fetos foram obtidos de fêmeas em cada idade gestacional (n = 3 YL; n = 4 Piau), e o perfil transcricional foi avaliado utilizando o microarranjo *Pigoligoarray* contendo 20.400 oligonucleotídeos. Um total de 486 oligonucleotídeos foram diferencialmente expressos (FC $\geq$ 1,5; FDR $\leq$ 0,05) entre 40 e 70 dias de gestação nos grupos YL ou Piau, e um total de 1.300 oligonucleotídeos foram diferencialmente expressos (FC $\geq$ 1,5; FDR $\leq$ 0,05) entre os suínos YL e Piau em ambas as idades gestacionais. De acordo com anotações propostas pelo Gene Ontology e por meio de análises de vias metabólicas, pôde-se determinar a classificação funcional dos genes diferencialmente expressos, e padrões de expressão raça-específico foram revelados. Treze genes foram selecionados para confirmar os resultados do microarranjo por meio de análises de qRT-PCR e os padrões de expressão da maioria dos genes foram confirmados, provendo informações relevantes sobre seus papéis desempenhados durante o processo do desenvolvimento muscular de suínos. Este estudo revelou padrões de expressão gênica tanto raça quanto idade-específicos durante o desenvolvimento fetal do músculo esquelético, incluindo genes até então não associados com o processo de miogênese. Tais informações podem contribuir para futuros avanços do melhoramento genético de suínos.

Palavras-chave: microarranjo, miogênese, suínos, qRT-PCR

Introduction

The most abundant tissue in animals is skeletal muscle, which typically accounts for 40 to 65% of the carcass weight in meat animal species. Development, growth and function of skeletal muscle are dynamic processes critical to animal survival, and involve the coordinated expression of thousands of genes. Development of skeletal muscle fibers during fetal growth in mammals takes place in two waves known as primary and secondary fiber formation (Wigmore & Evans, 2002). During each wave, myoblasts proliferate and fuse to form new muscle fibers. Primary fibers are formed *de novo*, whereas secondary fibers form around a primary fiber. In pigs, this process takes place at approximately 30-60 days (d) of gestation and 54-90 d of gestation for primary and secondary fibers, respectively (Wigmore & Stickland, 1983).

Several previous studies have evaluated gene expression patterns during pig fetal or embryonic development (Yelich *et al.* 1997; Wilson *et al.* 2000; Wesolowski *et al.* 2004). In addition, several studies have examined transcriptional profiles during various stages of pig skeletal muscle development in different breeds (Zhao *et al.* 2003; Lin & Hsu 2005; te Pas *et al.* 2005; Cagnazzo *et al.* 2006; Muráni *et al.* 2007; Tang *et al.* 2007; Li *et al.* 2008; Lobjois *et al.* 2008). It appears from these studies that there are breed-specific patterns of gene expression in developing pig skeletal muscle. In addition, as transcriptional profiling resources improve for exploring global gene expression patterns in the pig, it will be possible to gain further insight into the genes that are expressed during primary and secondary myofiber formation in pigs.

Piau is a naturalized Brazilian breed and genetic diversity exists within its current population (Sollero *et al.* 2009). Peixoto *et al.* (2006) have described this breed. The high fat content of Piau carcasses can be desirable in crossbreeding programs for improving meat quality traits such as marbling. Historically, this breed was used in the 1940s and 1950s in some genetic improvement programs due to its fat and meat characteristics (Vianna, 1956). Piau pigs have also recently been used for development of a resource population to identify quantitative trait loci for growth, carcass merit, meat quality and

reproductive traits (Peixoto *et al.* 2006; Paixão *et al.* 2008; Silva *et al.* 2008). However, investigation of Piau pigs at the transcriptome level has not been explored.

In this study, we use a whole-genome microarray to identify differentially expressed genes in *longissimus dorsi* (LD) muscle of Piau and Yorkshire-Landrace (YL) crossbred pigs at 40 and 70 d of gestation. The two developmental ages examined encompass the two waves of primary and secondary myofiber formation in pigs, and the two genetic groups allow comparison of breeds with distinctly different muscularity. Differentially expressed genes were functionally characterized, and selected genes are confirmed by quantitative real-time PCR analyses to increase understanding of the myogenesis process in pigs.

Material and methods

Animals and tissue sampling

Animal handling was done in accordance with regulations approved by the institutional animal welfare and ethics/protection commission of the *Universidade Federal de Viçosa* (UFV; DVT-UFV 02/2008) or by the Michigan State University (MSU) Institutional Animal Care and Use Committee (11/04-141-00). At the UFV Pig Breeding Farm, pregnant Piau gilts at either 40 d of gestation (n = 4) or 70 d of gestation (n = 4) 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). *Longissims dorsi* (LD) muscle samples were isolated from fetuses and placed in sterile tubes containing RNA*later*® (Qiagen). Samples were stored at 4ºC overnight and at -80ºC prior to RNA isolation. Pig fetuses were also collected from Yorkshire x Landrace (YL) crossbred gilts at 40 (n=3) and 70 (n=3) d of gestation by slaughtering gilts in the federally inspected MSU Meats Laboratory. Samples of LD muscle were obtained from fetuses, flash frozen in liquid nitrogen and stored at -80ºC.

RNA isolation and preparation of fluorescently labeled aRNA

Total RNA from approximately 30 mg RNA*later*®-stabilized LD tissue from Piau fetuses was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Total RNA from three Piau fetuses per litter was pooled for use in subsequent assays. Total RNA from approximately 1.0 g of LD tissue pooled from three fetuses from each of the YL gilts was extracted using TRIzol® reagent (Invitrogen Corp.) according to the manufacturer's instructions. Total RNA concentrations were measured using a NanoDrop spectrophotometer (NanoDrop Technologies, Inc.), and quality and integrity were determined with an RNA 6000 Labchip® kit using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.). Total RNA (1 µg) was reverse transcribed with an oligo(dT) primer and ArrayScript™ following the Amino Allyl MessageAmp™ II aRNA Kit (Ambion, Inc.) protocol. After

purification, 20 µg of each allyl-modified aRNA were dye coupled using NHS ester Cy™3 and Cy™5 dyes (GE Healthcare).

Oligonucleotide microarray hybridization and image processing

Transcriptional profiling was performed using the Swine Protein-Annotated Oligonucleotide Microarray (Pigoligoarray; www.pigoligoarray.org), which includes 20,400 70-mer oligonucleotides. We have previously evaluated this microarray for use in pig functional genomics studies and reported on the high quality of the array annotation (Steibel *et al.* 2009b). Hybridization and washing procedures were performed as previously described by our group (Steibel *et al*. 2009b). Images were processed at two different wavelengths (Cy3 and Cy5 fluorescence) to detect relative transcript abundance using an Axon GenePix® 4000B scanner (Molecular Devices) and the Gene Pix Pro 6.0 software (Molecular Devices). After spot alignment, non-normalized and non-background corrected median intensity values of each dye channel as well as normalized Cy5/Cy3 log ratios were stored as comma-separated values for analysis, and data were submitted to the National Center for Biotechnology Information's Gene Expression Omnibus database [GEO: GSE21412].

Experimental design and statistical analysis

In order to study the joint effects of age and breed, a factorial design was used to allow investigation of interactions between both factors. Gene expression patterns across ages and breeds were evaluated using a connected loop design (Tempelman 2005) with 13 microarray slides. Dyes were balanced so that each RNA sample was measured once with each dye.

Microarray data was normalized in the LIMMA software (Dudoit & Yang 2003) using a within-array global loess normalization (Yang *et al*. 2002). The Cy3 and Cy5 log normalized intensities were derived from the loess-normalized log Cy5/Cy3 intensities and analyzed using a linear mixed model (SAS v. 9.1.3):

$Y_{ijklm} = \mu + Dye_i + Breed_j + Age_k + BreedxAge_{jk} + Sample(BreedxAge)_{l:jk} + Array_m + e_{ijklm}$,

where Y_{ijklm} denotes the normalized log intensity for the l^{th} sample within the j^{th} breed and k^{th} age and labeled with the i^{th} dye, μ is the overall mean, Dye_i is the fixed effect of the i^{th} dye (i = Cy3, Cy5), $Breed_j$ is the fixed effect of the j^{th} breed (j = YL, Piau), Age_k is the fixed effect of the k^{th} age (k = 40, 70), $BreedxAge_{jk}$ pertains to the fixed interaction effects of the j^{th} breed with the k^{th} age, $Sample(BreedxAge)_{l:jk}$ is the random effect of the l^{th} sample (l = 1, 2, 3 for both ages of YL; l = 1, 2, 3, 4 for both ages of Piau) within the j^{th} breed and k^{th} age, $Array_m$ is the effect of the m^{th} array (m = 1, 2 …, 13), and e_{ijklm} is the residual. The software SAS® PROC MIXED was used for this analysis (Littell et al. 2006). In order to limit the occurrence of false positives, P-values were adjusted using the false discovery rate (FDR) procedure outlined by Storey and Tibshirani (2003).

Genes identified to be differentially expressed (FDR $\leq$ 0.05) between developmental ages or between breed types were evaluated using the DAVID software (Database for Annotation, Visualization and Integrated Discovery; http://david.abcc.ncifcrf.gov/; Dennis *et al*. 2003; Huang *et al.* 2009) in order to determine functional classifications based on gene ontology categories. Differentially expressed genes were further evaluated using the Ingenuity Pathways Analysis software (Ingenuity Systems) to identify biological pathways.

Real-time quantitative PCR (qRT-PCR) analysis

Quantitative real-time PCR (qRT-PCR) analyses using TaqMan® technology (Applied Biosystems) were used to confirm microarray results. Thirteen genes observed to be differentially expressed either between ages or between breeds were chosen for confirmation: *carbonic anhydrase III muscle specific* (*CA3*), *catenin (cadherin-associated protein) beta 1* (*CTNNB1*), *cathepsin L2* (*CTSL2*), *delta-like 1 homolog* (*DLK1*), *F-box protein 32* (*FBXO32*), *myozenin 1* (*MYOZ1*), *nebulin-related anchoring protein* (*NRAP*), *ornithine decarboxylase 1* (*ODC1*), *sarcopilin* (*SLN*), *signal transducer and activator of transcription 1* (*STAT1*), *TIMP metallopeptidase inhibitor 3* (*TIMP3*), *tenascin C* (*TNC*), and

ubiquitin specific peptidase 13 (isopeptidase T-3) (USP13). Each of the thirteen assays consisted of two unlabeled PCR primers and a FAM™ dye-labeled TaqMan® MGB (minor groove binder) probe. Ten of the primer-probe sets were commercially available from Applied Biosystems, whereas the remaining three (*USP13, ODC1* and *TIMP3*) were custom designed by Applied Biosystems for this project (Custom TaqMan® Gene Expression Assay). *Hypoxanthine phosphoribosyltransferase 1* (*HPRT1*; Von der Hardt *et al*., 2009; Lobjois *et al*., 2008) was used as a control gene for these analyses, and transcript abundance of *HPRT1* was consistent across all samples. TaqMan® primer-probe information is provided in Table S1. Total RNA (2 µg) from LD muscle samples (n=14) was reverse transcribed using random primers with the High Capacity cDNA Kit (Applied Biosystems) according to the manufacturer's instructions. Assays were performed in triplicate using 50 or 90 ng cDNA and the TaqMan 2X Universal PCR Master Mix (20 µL final volume per reaction) in an ABI Prism 7500 sequence detection system (Applied Biosystems). Cycling conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Transcription levels for each gene were normalized using the *HPRT1* gene. The cycles to threshold (C_t) data were analyzed using the %QPCR_MIXED macro developed in SAS (v. 9.1.3) based on a linear mixed model (Steibel at al., 2009a) and applied in this case for a completely randomized design with two factors.

Results

Identification of differentially expressed genes

A total of 1,043 oligonucleotides with HUGO Gene Nomenclature Committee (HGNC) annotation were found to be differentially expressed (FDR $\leq$ 0.05) in LD muscle of fetuses between 40 and 70 d gestation in either YL or Piau pigs. Only 268 of these oligos were differentially expressed between 40 and 70 d gestation fetuses in both breed types, whereas 261 were significant only in the YL pigs and 514 were significant only in the Piau pigs. Limiting the number of significantly differentially expressed oligos (FDR $\leq$ 0.05) to only those exhibiting FC differences $\geq$ 1.5 to help ensure reproducibility

(Shi *et al.* 2008) resulted in 197 oligos differentially expressed between 40 and 70 d of gestation in both breed types, 85 oligos differentially expressed in YL pigs only, and 204 differentially expressed in Piau pigs only (data not shown).

A total of 4,112 oligos were differentially expressed (FDR $\leq$ 0.05) between YL and Piau breed types at either 40 or 70 d of gestation. Of these, 1,728 oligos were differentially expressed between the two breed types at both 40 and 70 d of gestation, 1,508 oligos were significant only at 40 d of gestation, and 876 were significant only at 70 d of gestation. Limiting the number of significantly differentially expressed oligos (FDR $\leq$ 0.05) to only those exhibiting FC differences $\geq$ 1.5 resulted in 840 oligos differentially expressed between YL and Piau at both ages, 282 oligos differentially expressed only at 40 d of gestation, and 178 differentially expressed only at 70 d of gestation (data not shown).

Gene ontology annotation

In order to gain insight into the biological functions of differentially expressed transcripts in LD muscle between 40 and 70 d of gestation in the YL and Piau pigs, gene ontology (GO) annotations were evaluated using the DAVID software (Dennis *et al*. 2003; Huang *et al.* 2009). Category enrichment information based on biological process (BP), cellular component (CC) and molecular function (MF) GO terms are included in Table S3. Reduced lists are provided in Tables 1, 2 and 3 for BP, CC and MF GO terms, respectively.

The GO annotation was performed by considering four sets of differentially expressed genes; genes differentially expressed between developmental ages for each breed type and genes differentially expressed between the two breed types at each developmental age. The most highly enriched BP categories for the age comparisons in both breeds were muscle system related (Table 1) including muscle system process, muscle contraction, cytoskeletal organization and muscle organ development. Interestingly, the BP categories of phosphate metabolic process, and phosphorus metabolic process included 22 genes for the age comparison in the YL pigs, but these categories had no differentially expressed genes for the age comparison in the Piau pigs. The most enriched BP categories for the breed

comparison at 40 d of gestation were regulation of transcription, and transcription. However, these categories included no differentially expressed genes for the age comparisons in either breed or for the breed comparison at 70 d of gestation.

The most enriched GO CC categories for both the age and breed comparisons were non-membrane bound organelle and cytoskeleton (Table 2). For MF, the most enriched terms for the age comparisons were cytoskeletal protein binding and structural molecule activity (Table 3). The MF category of calcium ion binding included 31 differentially expressed genes in the age comparison for the Piau pigs; however there were no differentially expressed genes in this category for the age comparison of the YL pigs. The most enriched MF categories for the breed comparisons included nucleotide binding and DNA binding. Thus, the GO annotation indicates that the differentially expressed genes identified in this study are categorized into functional categories that would be expected for fetal LD muscle. In addition, several GO categories were identified to be enriched with differentially expressed genes for one breed type but not for the other, suggesting breed-specific developmental gene expression patterns.

Pathway analysis

The 467 genes observed to be differentially expressed in LD muscle between 40 and 70 d of gestation in either breed type (FDR $\leq$ 0.05; FC $\geq$ 1.5) were further functionally evaluated using the Ingenuity Pathways Analysis (IPA) software. This gene set was categorized into a total of 136 networks by the IPA software. A network containing 54 genes related to skeletal and muscular system development was observed for genes that were differentially expressed between 40 and 70 d gestation in both the YL and Piau breed types (Fig. 1). In this network, 39 genes were more highly expressed at 70 d of gestation (green color), whereas 15 genes were more highly expressed at 40 d of gestation (red color). While the same genes were observed to exhibit increased or decreased mRNA abundance in both the Piau and YL breed types in this network, the FC differences varied for some genes as depicted by the intensity of the green and red colors between panels (a) and (b) of Fig. 1. Interestingly, several genes exhibit higher expression at 70 d in the Piau pigs (higher green intensity), while several genes exhibit higher

expression at 40 d in the YL pigs (higher red intensity) suggesting the potential for a breed-specific expression pattern.

An additional network that included differentially expressed genes between 40 and 70 d gestation in both Piau and YL pigs was related to cellular function and maintenance (data not shown). For this network, 10 genes were more highly expressed at 70 d of gestation and 4 genes were more highly expressed at 40 d of gestation. A third informative network, cellular morphology and cellular assembly, involved 53 genes that were differentially expressed between the two ages only in the Piau breed (data not shown). Thirty-five of these genes exhibited increased expression at 70 d of gestation, whereas 18 exhibited higher expression at 40 d of gestation.

Quantitative RT-PCR confirmation of differentially expressed genes

A total of 13 differentially expressed genes were selected for confirmation by qRT-PCR. Selected genes represented various functional groups including myofibrillar genes (*MYOZ1* and *NRAP*), proliferation or differentiation (*DLK1* and *ODC1*), metabolic processes (*SLN* and *CA3*), extracellular matrix (*TIMP3* and *TNC*), signal transduction or transcription activation (*STAT1* and *CTNNB1*), and phosphorylation-dependent ubiquitination process or proteolysis mechanisms (*CTSL2*, *FBXO32* and *USP13*). Also, five of these genes (*DLK1*, *NRAP*, *MYOZ1*, *SLN* and *TNC*) were included in the skeletal and muscular system development network shown in Fig. 1.

Nine of the selected genes exhibited differential expression in LD between 40 and 70 d of gestation in at least one breed type using the Pigoligoarray, and expression patterns for all of these genes were confirmed by qRT-PCR ($P \leq 0.01$; Fig. 2). *CA3*, *DLK1*, *MYOZ1*, *NRAP* and *SLN* all exhibited higher mRNA abundance at 70 d of gestation in both the YL and Piau breed types on the microarray, and these expression patterns were confirmed by qRT-PCR. Expression of *TNC* was observed to be higher in LD at 40 d of gestation in both YL and Piau pigs on the microarray and this was also confirmed by qRT-PCR. Microarray results for *CTSL2* indicated higher expression at 40 d of gestation for Piau pigs, but no difference between ages for YL pigs, and this result was confirmed with the qRT-PCR assay. *FBXO32*

and *USP13* exhibited higher expression at 70 d of gestation in Piau pigs on the microarray, but no significant difference was observed for the YL pigs using FDR ≤ 0.05. Results of the qRT-PCR analyses for these genes confirmed that expression was higher in LD at 70 d for the Piau pigs, and also indicated expression to be higher at 70 d for the YL pigs.

Results for qRT-PCR analyses of six genes significantly differentially expressed by microarray analysis in LD at either 40 or 70 d of gestation or at both ages when comparing YL and Piau pigs are shown in Fig. 3. This gene set includes two of the genes that were also differentially expressed between ages (*FBXO32* and *NRAP*). Differential expression of most of these genes was confirmed. Expression of *NRAP* was observed to be higher in LD of Piau pigs at both 40 and 70 d of gestation on the microarray and this result was confirmed by qRT-PCR. *FBXO32* exhibited significantly higher expression in LD from Piau pigs than in LD from YL pigs at 70 d of gestation, but no difference in expression between breed types was observed at 40 d of gestation, and this result was also confirmed by qRT-PCR. Results for *ODC1* on the microarray indicated that expression was higher in Piau pigs at 70 d of gestation, but no breed type differences were observed at 40 d of gestation with an FDR ≤ 0.05. Results for *ODC1* using qRT-PCR confirmed higher expression in Piau pigs at 70 d of gestation and also indicated that expression was significantly higher in Piau pigs at 40 d of gestation. Similarly, microarray results for *TIMP3* indicated that *TIMP3* expression was higher in Piau pigs at 40 d of gestation but breed type differences were not significant at 70 d of gestation (FDR ≤ 0.05). Results for qRT-PCR analysis of *TIMP3* revealed significantly higher expression of *TIMP3* in LD of Piau pigs at both 40 d and 70 d of gestation. On the microarray, *CTNNB1* exhibited a 4.1-fold higher expression in YL pigs than Piau pigs at 40 d of gestation with no breed difference observed at 70 d of gestation. There was also no difference in expression between breed types for this gene at 70 d observed by qRT-PCR analysis. However, qRT-PCR failed to confirm the breed difference that was seen with the microarray in the 40 d samples. *STAT1* exhibited significantly higher expression in Piau pigs at both 40 and 70 d of gestation by qRT-PCR analysis. However, this result was opposite to the *STAT1* expression pattern observed using the Pigoligoarray. On

both platforms, transcript abundance levels for *CTNNB1* and *STAT1* were relatively low which could account for the inconsistent results for these genes.

Discussion

This study evaluated transcriptional profiles in LD muscle of pigs at 40 and 70 d of gestation, two developmental stages that encompass the primary and secondary waves of muscle fiber formation in pigs (Wigmore & Stickland, 1983; Wigmore & Evans, 2002). Samples obtained from Yorkshire-Landrace (YL) crossbred pigs and Piau pigs (a native Brazilian breed), breed types that differ in muscularity, were used in order to allow breed type comparisons. Our results revealed a large number of differentially expressed genes both between developmental ages and between breed types. In general, the developmental transcript profiles for the Piau and YL pigs were similar, although breed-specific patterns of gene expression were revealed. In addition, the relative abundance of transcripts (based on fluorescence intensity using the microarray) tended to be greater for the Piau pigs at 70 d of gestation suggesting that gene expression in LD muscle of YL pigs may be more delayed than in Piau pigs. This observation is consistent with results reported by Cagnazzo *et al*. (2006) who reported a study examining gene expression in developing skeletal muscle of Duroc and Pietrain pigs, breeds that differ for muscle fiber characteristics as well as growth and muscularity phenotypes. This group observed differential expression for numerous myogenesis-related genes and suggested that pigs of the heavier muscled Pietrain breed may exhibit a delayed myogenesis process, perhaps resulting in increased numbers of myofibers, relative to pigs of the Duroc breed.

Following global transcript profiling analysis using the microarray, several genes were selected for confirmation by qRT-PCR. Some of these genes encode products with functions related to skeletal muscle structure or function. Myozenin proteins including *MYOZ1* appear to serve as intracellular binding proteins involved in linking Z-disk proteins such as *alpha-actinin* and *titin-cap*, and *MYOZ1* plays a complex role in the modulation of calcineurin signaling (Faulkner *et al*., 2000). We observed significantly higher expression of *MYOZ1* at 70 d of gestation relative to 40 d of gestation in both Piau

and YL pigs. This is consistent with recent observations by Raymond *et al.* (2010) who compared gene expression patterns in human skeletal muscle tissue and cultured myotubes, and found *MYOZ1* levels to be significantly downregulated in cultured myotubes supporting the role of this gene in tissue structure and maturation. The product of the NRAP gene is associated with developing skeletal muscle myofibrillar structures (Lu *et al.* 2008). Our results indicated that NRAP expression was higher at 70 d of gestation in both Piau and YL pigs, and also expression was higher in Piau pigs at both ages. Similar results were reported by Muráni *et al*. (2007) who observed higher expression of NRAP in skeletal muscle of Pietrain fetuses at 35 d of gestation as compared to Duroc fetuses. The *SLN* gene encodes a small proteolipid that regulates several sarcoplasmic reticulum Ca(2+)-ATPases. Our results indicated that *SLN* expression was higher at 70 d of gestation in both Piau and YL pigs, with highest expression in the Piau pigs. This agrees with results reported by Lin and Hsu (2005), who compared neonatal pigs of a native breed (Taoyuan) to Duroc pigs and observed higher expression in the Taoyuan pigs.

Additional genes selected for qRT-PCR confirmation encode products involved with cellular or tissue growth. The tissue inhibitors of metalloproteinsases (TIMP) were originally characterized as inhibitors of matrix metalloproteinases, but it is now known that they have a much wider range of biological activities including effects on cell growth and differentiation, cell migration and apoptosis among others (Brew & Nagase, 2010). We observed higher expression of *TIMP3* in Piau pigs vs. YL pigs at both 40 and 70 d of gestation indicating a breed-specific expression pattern for this gene. Similarly, we also observed higher expression of *ODC1* in Piau pigs at both 40 and 70 d of gestation. The *ODC1* gene encodes the rate-limiting enzyme in the polyamine biosynthesis pathway. MacLean *et al*. (2008) found *ODC1* expression to be decreased in skeletal muscle of androgen receptor knockout mice, and based on their other microarray results and observations of muscle mass they concluded that androgens promote muscle growth by maintaining myoblasts in the proliferative state and by delaying differentiation. While it is not possible to determine if hormonal influences are involved in the breed-specific gene expression patterns that we observed, these observations taken together with the potential delay in gene expression for the YL pigs in this study and the Duroc pigs in the Cagnazzo *et al*. (2006)

study may support the possibility of a prolonged proliferative period which could result in increased numbers of primary and secondary fibers as suggested by Cagnazzo et al. (2006).

The *CA3* gene is the muscle-specific member of a multigene family encoding metalloenzymes that catalyse the reversible hydration of carbon dioxide. Our results indicated that *CA3* expression was higher at 70 d of gestation than at 40 d of gestation in both breed types, which is in agreement with the results of Wang *et al*. (2006) who reported expression of CA3 to increase in skeletal muscle from 33 to 65 d of gestation in Tongcheng pigs. The *DLK1* gene was observed to exhibit higher expression at 70 d of gestation in both breed types. This gene, located in an imprinted region of mammalian genomes, is involved in differentiation and it has been shown to be up-regulated in skeletal muscle of callipyge lambs that exhibit extreme muscle hypertrophy (Fleming-Waddell *et al*. 2007).

The *TNC* gene product is a mechano-regulated, morphogenic, extracellular matrix protein that is associated with tissue remodeling (Flück *et al*. 2008) and it is involved in innervation during development. We observed *TNC* to be more highly expressed at 40 d of gestation in both Piau and YL pigs, suggesting a greater role for *TNC* during primary fiber formation.

Two genes selected for qRT-PCR analysis did not confirm the results observed with the microarray. The *CTNNB1* gene product is a primary mediator of the WNT/β-catenin signaling pathway that when activated, leads to the stabilization of β-catenin which enters the nucleus to activate target genes including *MYOD* and *MYF5* (Armstrong & Esser, 2005) potentially increasing myoblast proliferation. Expression of *CTNNB1* was significantly higher in YL pigs vs. Piau pigs on the microarray at 40 d of gestation, although no breed difference was observed at 70 d of gestation. Using qRT-PCR, no significant breed difference was observed for this gene at either age. We detected relatively low signal intensities for this gene on both platforms which could account for the inconsistent results. Cagnazzo et al. (2006) observed *CTNNB1* to have increased expression in Pietrain fetuses at several developmental ages from 35 to 91 d of gestation compared to Duroc fetuses. Thus, this gene may exhibit breed-specific expression patterns and further study may be warranted. The *STAT1* gene encodes a protein that is phosphorylated and acts as a transcriptional activator through a signal transduction pathway mediated by

interferons. Sun *et al.* (2007) reported that *STAT1* was an essential part of a *JAK1-STAT1-STAT3* signaling pathway that promotes myoblast proliferation, supporting a role for this gene in fetal muscle development. On the microarray, *STAT1* exhibited significantly higher expression in LD muscle of YL pigs. However, results of the qRT-PCR assay indicated that expression of this gene was significantly higher in the Piau pigs. Low signal intensities on both platforms may have contributed to the inconsistent results.

Three genes selected for qRT-PCR confirmation are known to function in proteolysis pathways (*CTSL2*, *FBXO32* and *USP13*). Ubiquitin specific proteases are enzymes that remove ubiquitin from specific protein substrates and allow protein salvage from proteosome degradation, or regulation of protein localization or activation. Little information is available regarding the specific function of *USP13*, however, the mRNA abundance of this gene is higher in skeletal muscle than in any other tissue in mice and humans (http://biogps.gnf.org). We observed higher expression of *USP13* in LD muscle at 70 d of gestation in both Piau and YL pigs suggesting involvement of this protein as secondary muscle fibers are formed. *CTSL2* encodes a lysosomal cysteine proteinase and this gene has been shown to be induced as a result of muscle atrophy (Lecker *et al.* 2004). We observed increased expression of *CTSL2* at 40 d of gestation in Piau pigs, but no differences in expression between ages for YL pigs. Thus, while additional studies are needed, this gene may be involved in protein turnover, and developmental regulation of gene expression may be breed-specific. *FBXO32* (also known as atrogin-1 and muscle atrophy F-box, MAFbx) is an F-box containing ubiquitin protein ligase that was identified in mice and rats due to its high level of expression during muscle atrophy (Bodine *et al.* 2001; Gomes *et al.* 2001). Tintignac *et al.* (2005) also reported a role for *FBXO32* in muscle differentiation by determining that *FBXO32* mediates degradation of *MYOD* via the ubiquitin proteosome pathway. We observed higher expression of *FBXO32* in LD at 70 d of gestation in both Piau and YL pigs, and also Piau pigs exhibited significantly higher expression of *FBXO32* than YL pigs at 70 d of gestation, suggesting that expression of this gene is developmentally regulated and breed-type specific. These results support a role for the ubiquitin proteosome system in fetal skeletal muscle development.

In summary, we have used a whole genome microarray (Pigoligoarray) to examine transcriptional profiles in LD muscle at 40 and 70 d of gestation in Piau and YL crossbred pigs. These gestational ages encompass the two waves of primary and secondary muscle fiber formation in pigs. We have identified both developmental and breed-specific patterns of gene expression. In addition, this study is the first report to evaluate gene expression in Brazilian native Piau pigs. Skeletal muscle development involves the synchronized expression and interaction of many genes, and this study provides additional insight into the process of myogenesis in pigs.

Acknowledgements

This project was supported by National Research Initiative Competitive Grant no. 2003-35206-13922 from the USDA National Institute of Food and Agriculture. The authors are grateful to the Brazilian funding agencies FAPEMIG, CAPES and CNPq. We thank M. Doumit and R. Bates for technical assistance. Support from the US Swine Genome Coordinator for synthesis of oligonucleotides for the Pigoligoarray is greatly appreciated.

Supporting information

Table S1 (Appendix 1) Primer information for genes selected for qRT-PCR.

Table 1 Over-represented gene ontology biological processes for differentially expressed genes in LD muscle of Piau and Yorkshire-Landrace (YL) pigs at 40 and 70 d of gestation.

	Piau (40 vs. 70 d)[1]		YL (40 vs. 70 d)[2]		40 d (YL vs. P)[3]		70 d (YL vs. P)[4]	
Biological Process GO Term (Gene Ontology ID No.)	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value
muscle system process (3012)	30	1.17E-18	22	5.52E-14			15	8.28E-02
muscle contraction (6936)	30	7.93E-20	22	8.20E-15			15	4.44E-02
cytoskeleton organization (7010)	25	1.75E-05	20	2.54E-05	41	7.31E-03	39	3.82E-03
muscle organ development (7517)	25	1.64E-11	13	7.12E-05			23	3.20E-03
response to organic substance (10033)	24	3.13E-02	22	2.31E-03	59	2.10E-02	50	7.93E-02
macromolecular complex subunit organization (43933)	24	2.67E-02	20	9.01E-03	69	1.62E-04	67	2.20E-05
homeostatic process (42592)	24	4.61E-02						
macromolecular complex assembly (65003)	23	2.43E-02	19	9.78E-03	66	1.27E-04	64	1.91E-05
cell adhesion (7155)	23	4.01E-02						
biological adhesion (22610)	23	4.02E-02						
protein complex assembly (6461)	20	1.07E-02	17	3.42E-03	46	7.71E-03	43	5.59E-03
protein complex biogenesis (70271)	20	1.07E-02	17	3.42E-03	46	7.71E-03	43	5.59E-03
generation of precursor metabolites and energy (6091)	20	3.62E-05	16	6.47E-05	33	3.25E-03	35	1.33E-04
cellular component morphogenesis (32989)	20	7.66E-04	14	6.19E-03	35	3.14E-02		
phosphate metabolic process (6796)			22	5.16E-02				
phosphorus metabolic process (6793)			22	5.16E-02				
regulation of transcription (45449)					184	2.22E-02		
transcription (6350)					149	3.84E-02		
protein localization (8104)					85	3.39E-05	79	2.22E-05
cell cycle (7049)					82	1.53E-06	71	3.16E-05
positive regulation of macromolecule metabolic process (10604)					81	1.03E-04	72	3.80E-04
proteolysis (6508)					79	4.79E-02	80	3.31E-03
macromolecule catabolic process (9057)					78	2.22E-05	74	6.35E-06
negative regulation of macromolecule metabolic process (10605)					76	8.23E-06	73	1.33E-06
cellular macromolecule catabolic process (44265)					74	1.81E-05	71	3.31E-06
protein transport (15031)					74	9.26E-05	70	3.29E-05
establishment of protein localization (45184)					74	1.23E-04	70	4.34E-05

[1]Differentially expressed genes in Piau pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table.

[2]Differentially expressed genes in YL pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table unless category included for previous contrast.

[3]Differentially expressed genes at 40 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.

[4]Differentially expressed genes at 70 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.

Table 2 Over-represented gene ontology cellular components for differentially expressed genes in LD muscle of Piau and Yorkshire-Landrace (YL) pigs at 40 and 70 d of gestation.

	Piau (40 vs. 70 d)[1]		YL (40 vs. 70 d)[2]		40 d (YL vs. P)[3]		70 d (YL vs. P)[4]	
Cellular Component GO Term (Gene Ontology ID No.)	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value
non-membrane-bounded organelle (43228)	70	4.87E-02	54	2.40E-02	222	2.27E-08	197	4.72E-07
intracellular non-membrane-bounded organelle (43232)	70	4.87E-02	54	2.40E-02	222	2.27E-08	197	4.72E-07
cytoskeleton (5856)	56	9.16E-06	41	1.04E-04	99	6.21E-02	91	4.73E-02
cytoskeletal part (44430)	36	1.95E-03	29	1.03E-03	72	4.42E-02	71	7.23E-03
cell fraction (267)	36	1.42E-02	24	8.90E-02				
extracellular region part (44421)	35	4.07E-03						
contractile fiber (43292)	34	2.86E-27	19	5.42E-13			15	6.59E-03
myofibril (30016)	32	5.12E-26	18	1.51E-12			14	7.84E-03
contractile fiber part (44449)	31	1.72E-24	19	1.60E-13			14	9.07E-03
sarcomere (30017)	29	6.79E-24	18	1.81E-13			14	2.69E-03
actin cytoskeleton (15629)	27	2.44E-10	19	2.61E-07	23	9.91E-02	28	1.83E-03
vesicle (31982)	22	6.60E-02	25	1.49E-04	54	3.08E-02		
cytoplasmic vesicle (31410)	21	7.59E-02	24	2.06E-04	50	6.25E-02		
extracellular matrix (31012)	21	8.11E-05			42	3.17E-05	29	2.53E-02
proteinaceous extracellular matrix (5578)	20	8.96E-05			42	5.13E-06	29	1.01E-02
cytosol (5829)			35	3.21E-03	141	3.59E-11	128	1.90E-10
organelle membrane (31090)			27	2.38E-02	87	8.92E-03		
endoplasmic reticulum (5783)			22	8.04E-02	93	7.94E-06	85	1.16E-05
cytoplasmic membrane-bounded vesicle (16023)			21	4.49E-04	48	1.21E-02	40	6.03E-02
endomembrane system (12505)			19	6.97E-02	71	7.28E-04	55	4.78E-02
organelle envelope (31967)			17	3.69E-02	53	1.20E-02	46	3.29E-02
envelope (31975)			17	3.78E-02	53	1.28E-02	46	3.50E-02
membrane-enclosed lumen (31974)					198	2.66E-16	177	2.79E-14
organelle lumen (43233)					195	2.95E-16	174	4.06E-14
intracellular organelle lumen (70013)					191	6.75E-16	169	2.03E-13
nuclear lumen (31981)					156	6.46E-13	134	9.42E-10
nucleoplasm (5654)					105	3.07E-11	88	4.92E-08
mitochondrion (5739)					83	2.57E-02	75	3.05E-02
Golgi apparatus (5794)					71	1.08E-02	66	6.86E-03

[1]Differentially expressed genes in Piau pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table.
[2]Differentially expressed genes in YL pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table unless category included for previous contrast.
[3]Differentially expressed genes at 40 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.
[4]Differentially expressed genes at 70 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.

Table 3 Over-represented gene ontology molecular functions for differentially expressed genes in LD muscle of Piau and Yorkshire-Landrace (YL) pigs at 40 and 70 d of gestation.

	Piau (40 vs. 70 d)[1]		YL (40 vs. 70 d)[2]		40 d (YL vs. P)[3]		70 d (YL vs. P)[4]	
Molecular Function GO Term (Gene Ontology ID No.)	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value	Gene Count	P-value
cytoskeletal protein binding (8092)	40	7.05E-13	24	2.61E-06	57	3.96E-06	53	2.60E-06
structural molecule activity (5198)	35	2.80E-07	23	2.69E-04				
calcium ion binding (5509)	31	7.96E-03						
actin binding (3779)	28	7.20E-10	18	1.04E-05	44	6.64E-07	39	3.70E-06
identical protein binding (42802)	23	1.28E-02	22	7.72E-04				
nucleotide binding (166)			48	1.23E-02	184	1.69E-06	168	9.36E-07
purine nucleotide binding (17076)			42	1.43E-02	138	1.01E-02	126	6.18E-03
ribonucleotide binding (32553)			42	6.97E-03	134	7.31E-03	122	4.97E-03
purine ribonucleotide binding (32555)			42	6.97E-03	134	7.31E-03	122	4.97E-03
nucleoside binding (1882)			35	3.04E-02			100	6.07E-02
purine nucleoside binding (1883)			34	4.33E-02			99	6.22E-02
adenyl ribonucleotide binding (32559)			33	3.01E-02	102	9.09E-02	92	8.28E-02
ATP binding (5524)			33	2.55E-02	101	8.75E-02		
adenyl nucleotide binding (30554)			33	5.53E-02			96	9.06E-02
DNA binding (3677)					165	8.66E-03	138	8.90E-02
transcription regulator activity (30528)					116	2.76E-03	98	2.36E-02
RNA binding (3723)					85	1.14E-09	71	3.97E-07

[1]Differentially expressed genes in Piau pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table.
[2]Differentially expressed genes in YL pigs between 40 and 70 d of gestation. Categories enriched with 20 or more genes included in table unless category included for previous contrast.
[3]Differentially expressed genes at 40 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.
[4]Differentially expressed genes at 70 d of gestation between YL and Piau pigs. Categories enriched with 70 or more genes included in table unless category included for previous contrast.

a)

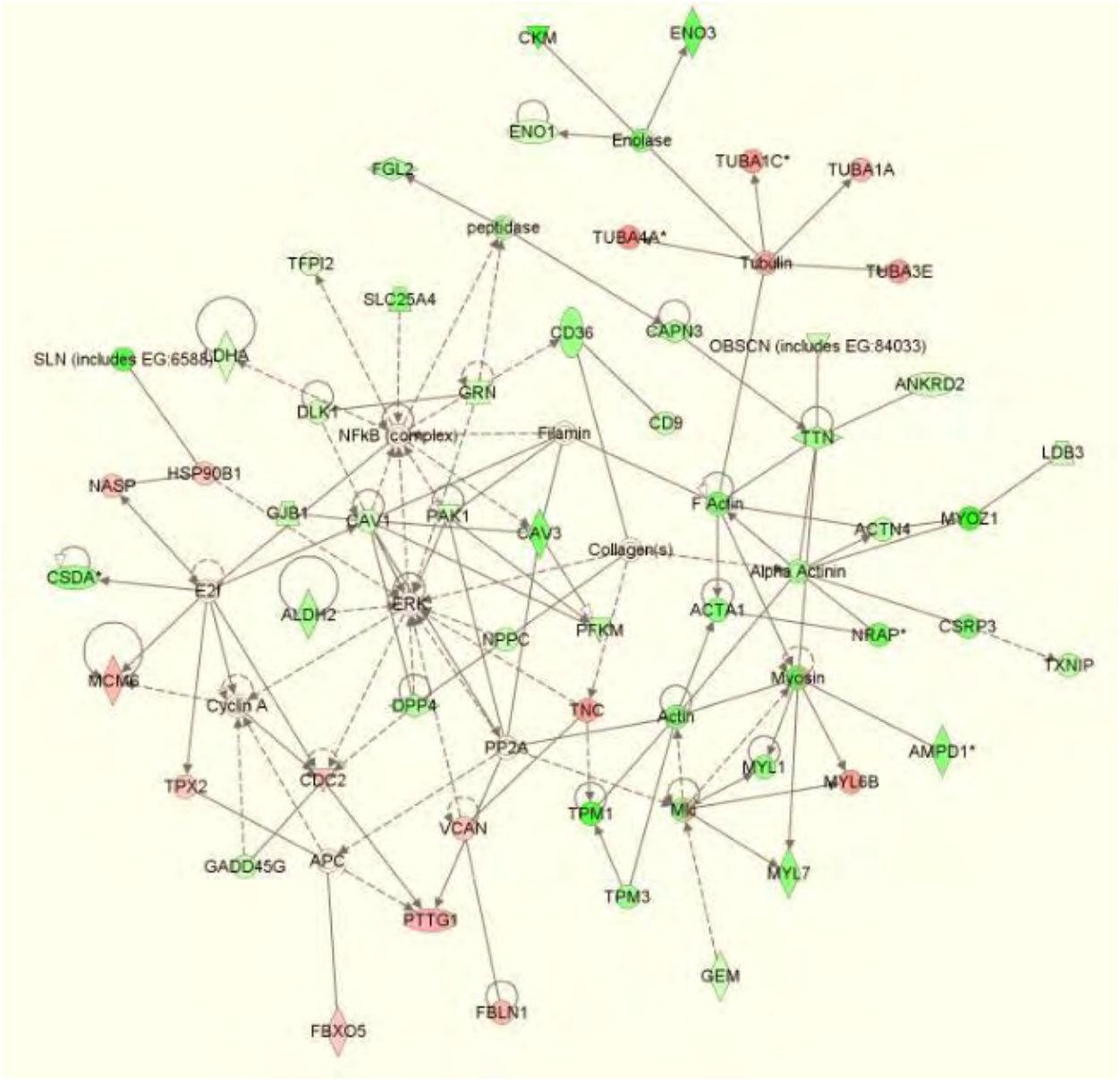


Figure 1 Skeletal and Muscular System Development and Function gene network containing 54 genes differentially expressed in LD muscle between 40 and 70 d of gestation derived using the Ingenuity Pathways Analysis System software (Ingenuity Systems®). (a) Differentially expressed genes in Piau pigs. (b) Differentially expressed genes in Yorkshire-Landrace (YL) crossbred pigs. Red, higher mRNA abundance in LD muscle at 40 d of gestation relative to 70 d of gestation; Green, higher mRNA abundance in LD muscle at 70 d of gestation relative to 40 d of gestation. Darker colors signify higher fold change expression differences.

b)

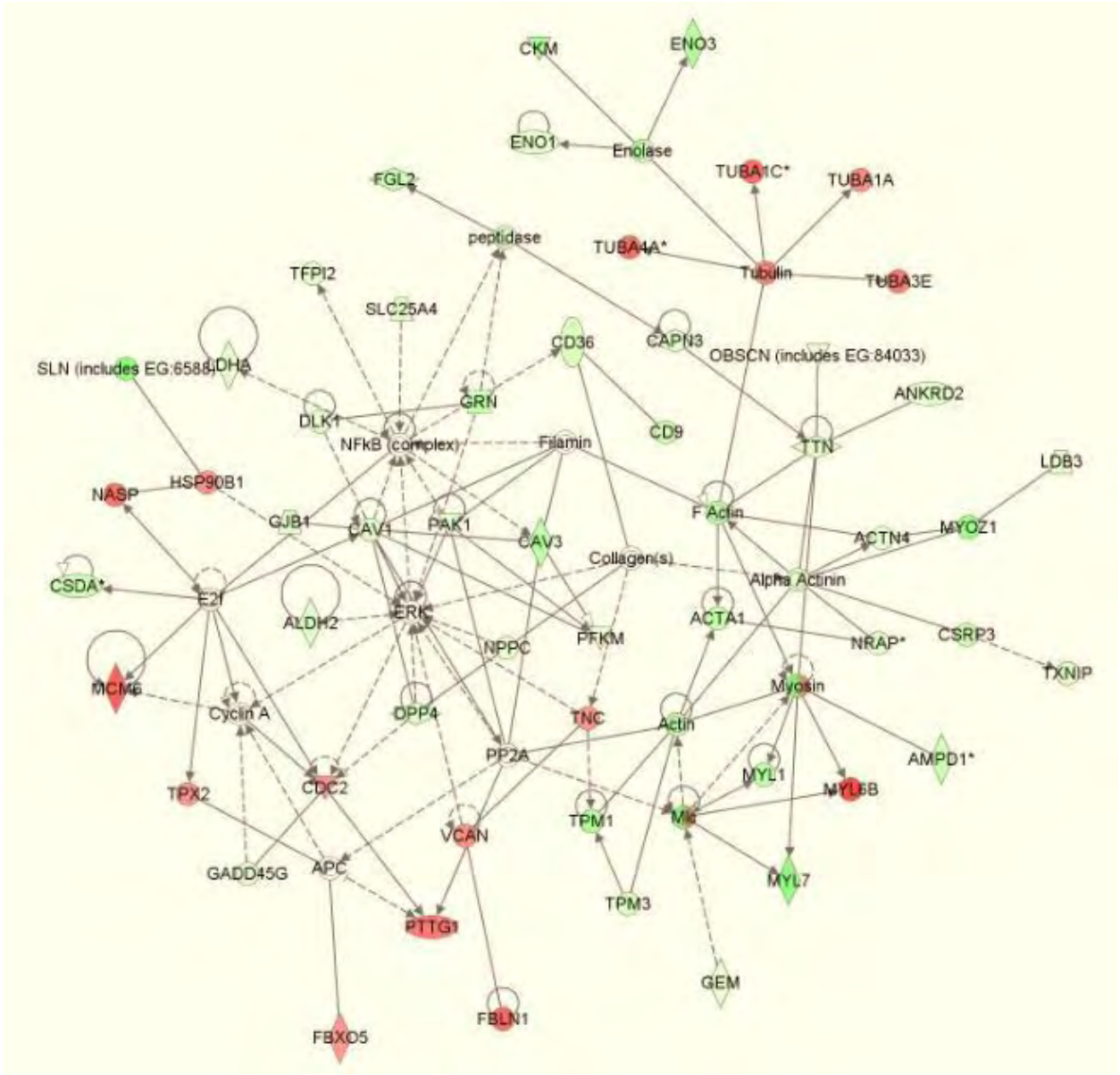


Figure 1 Skeletal and Muscular System Development and Function gene network containing 54 genes differentially expressed in LD muscle between 40 and 70 d of gestation derived using the Ingenuity Pathways Analysis System software (Ingenuity Systems®). (a) Differentially expressed genes in Piau pigs. (b) Differentially expressed genes in Yorkshire-Landrace (YL) crossbred pigs. Red, higher mRNA abundance in LD muscle at 40 d of gestation relative to 70 d of gestation; Green, higher mRNA abundance in LD muscle at 70 d of gestation relative to 40 d of gestation. Darker colors signify higher fold change expression differences.

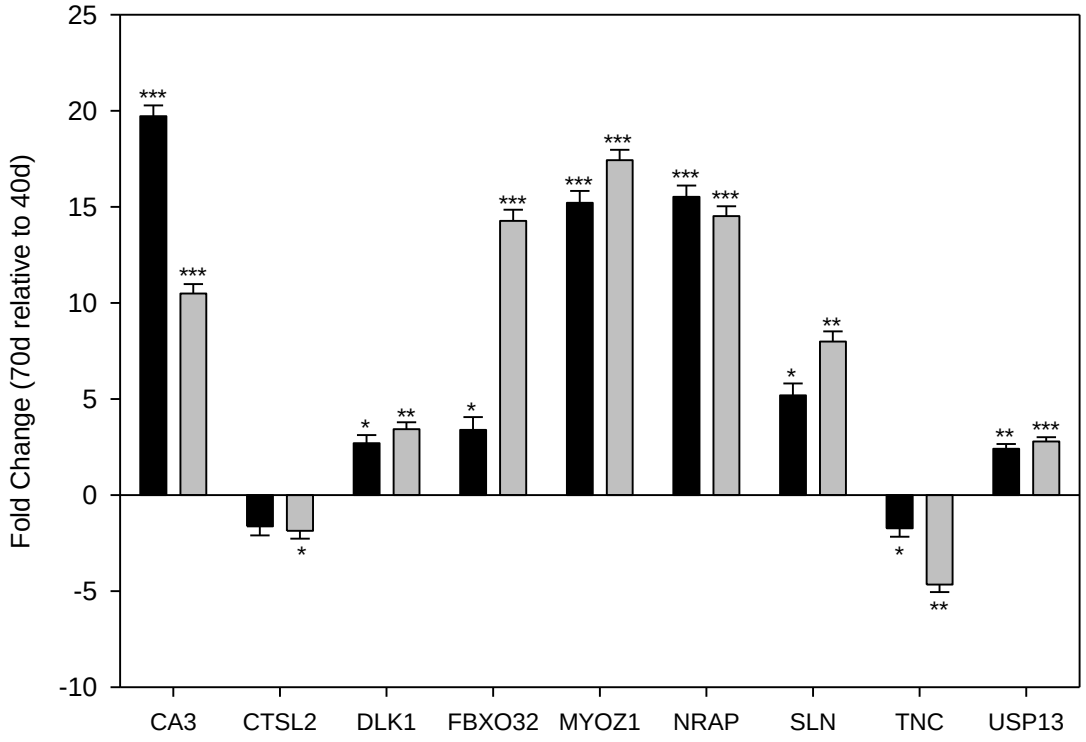


Figure 2 Quantitative RT-PCR results for nine genes observed by microarray analysis to exhibit differential expression in LD muscle between 40 and 70 d of gestation. Results are presented as fold changes for expression at 70 d of gestation relative to expression at 40 d of gestation such that bars above the origin indicate higher expression at 70 d and bars below the origin indicate lower expression at 70 d. Black bars, Yorkshire-Landrace (YL) pigs; Gray bars, Piau pigs. * $P \leq 0.01$; ** $P \leq 0.001$; *** $P \leq 0.0001$.

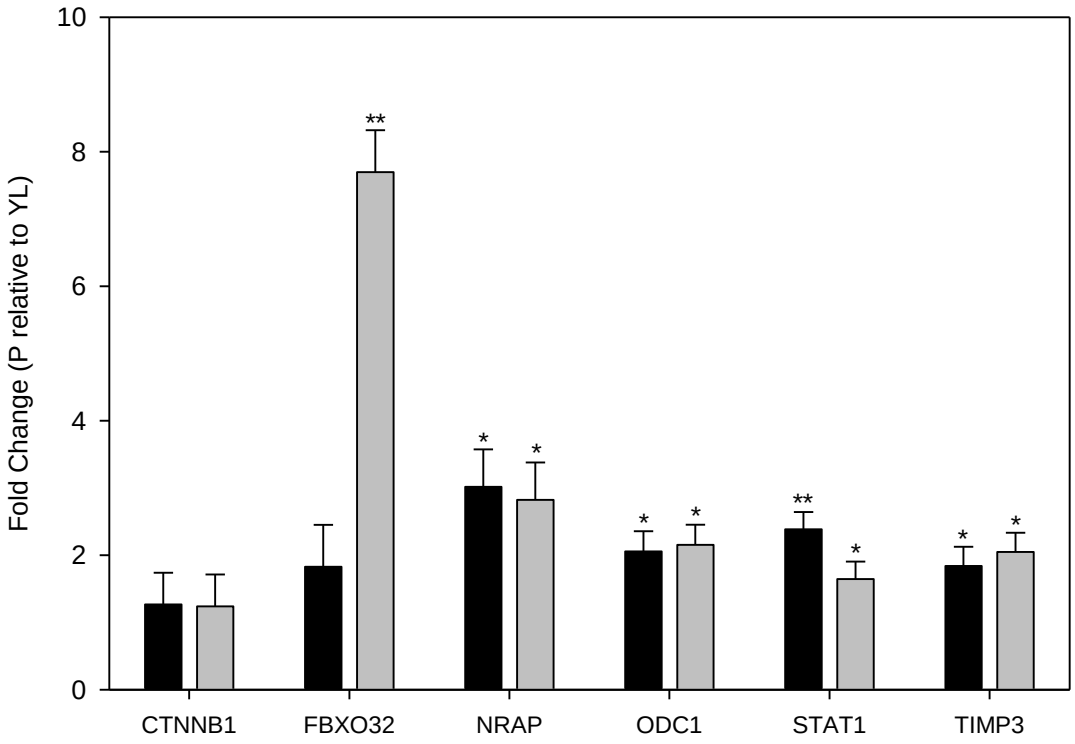


Figure 3 Quantitative RT-PCR results for six genes observed by microarray analysis to exhibit differential expression in fetal LD muscle between Yorkshire-Landrace (YL) and Piau pigs. Results are presented as fold changes for expression in Piau pigs relative to expression in YL pigs such that bars above the origin indicate higher expression in Piau pigs. Black bars, 40 d of gestation; Gray bars, 70 d of gestation. * $P \leq 0.01$; ** $P \leq 0.001$.

References

Armstrong D.D. & Esser K.A. (2005) Wnt/β-catenin signaling activates growth-control genes during overload-induced skeletal muscle hypertrophy. *American Journal of Physiology Cell Physiology* **289**, 853-59.

Bodine S.C., Latres E., Baumhueter S., Lai V.K.-M., Nunez L., Clarke B.A., Poueymirou W.T., Panaro F.J., Na E., Dharmarajan K., Pan Z.-Q., Valenzuela D.M., DeChiara T.M., Stitt T.N., Yancopoulos G.D. & Glass D.J. (2001) Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science* **294**, 1704-8.

Brew K. & Nagase H. (2010) The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. *Biochimica and Biophysica Acta* **1803**, 55-71.

Cagnazzo M., te Pas M.F.W., Priem J., de Wit A.A.C., Pool M.H., Davoli R. & Russo V. (2006) Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. *Journal of Animal Science* **84**, 1-10.

Dennis G. Jr., Sherman B.T., Hosack D.A., Yang J., Gao W., Lane H.C. & Lempicki R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* **4**, P3.

Du M., Yan X., Tong J.F., Zhao J. & Zhu M.J. (2010) Maternal Obesity, Inflammation, and Fetal Skeletal Muscle Development. *Biology of Reproduction* **82**, 4-12.

Dudoit S. & Yang Y.H. (2003) Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data. In: *The Analysis of Gene Expression Data: Methods and Software* (ed. by G. Parmigiani, E.S. Garrett, R.A. Irizarry & S.L. Zeger), pp. 73-101, Springer, New York.

Faulkner G., Pallavicini A., Comelli A., Salamon M., Bortoletto G., Ievolella C., Trevisan S., Kojic S., Dalla Vecchia F., Laveder P., Valle G. & Lanfranchi G. (2000) FATZ, a filamin-, actinin-, and telethonin-binding protein of the Z-disc of skeletal muscle. *Journal of Biological Chemistry* **275**, 41234-242.

Fleming-Waddell J.N., Wilson L.M., Olbricht G.R., Vuocolo T., Byrne K., Craig B.A., Tellam R.L., Cockett N.E. & Bidwell C.A. (2007) Analysis of gene expression during the onset of muscle hypertrophy in callipyge lambs. *Animal Genetics* **38**, 28-36.

Flück M., Mund S.I., Schittny J.C., Klossner S., Durieux A. & Giraud M.L. (2008) Mechano-regulated Tenascin-C orchestrates muscle repair. *Proceedings of the National Academy of Sciences* **105**, 13662-667.

Gomes M.D., Lecker S.H., Jagoe R.T., Navon A. & Goldberg A.L. (2001) Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. *Proceedings of the National Academy of Sciences* **98**, 14440-5.

Huang da W., Sherman B.T. & Lempicki R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57.

Lecker S.H., Jagoe R.T., Gilbert A., Gomes M., Baracos V., Bailey J., Price S.R., Mitch W.E. & Goldberg A.L. (2004) Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *The FASEB Journal* **18**, 39-51.

Li M., Li X., Zhu L., Teng X., Xiao H., Shuai S., Chen L., Li Q. & Guo Y. (2008) Differential expression analysis and regulatory network reconstruction for genes associated with muscle growth and adipose deposition in obese and lean pigs. *Natural Science* **18**, 387-99.

Lin C.S. & Hsu C.W. (2005) Differentially transcribed genes in skeletal muscle of Duroc and Taoyuan pigs. *Journal of Animal Science* **83**, 2075-86.

Littell R.C., Milliken G.A., Stroup W.W., Wolfinger R.D. & Schabenberger O. (2006) SAS for Mixed Models, Second edition. SAS Institute Inc., Cary, NC, USA.

Lobjois V., Liaubet L., SanCristobal M., Glenisson J., Feve K., Rallieres J., Le Roy P., Milan D., Cherel P. & Hatey F. (2008) A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *Animal Genetics* **39**, 147-62.

Lu S., Borst D.E. & Borowits R. (2008) Expression and alternative splicing of N-RAP during mouse skeletal muscle development. *Cell Motility and the Cytoskeleton* **65**, 945-54.

MacLean H.E., Chiu W.S.M., Notini A.J., Axell A-M., Davey R.A, McManus J.F., Ma C., Plant D.R., Lynch G.S. & Zajac J.D. (2008) Impaired skeletal muscle development and function in male, but not female, genomic androgen receptor knockout mice. *The FASEB Journal* **22**, 2676-89.

Muráni E., Murániová M., Ponsuksili S., Schellander K. & Wimmers K. (2007) Identification of genes differentially expressed during prenatal development of skeletal muscle in two pig breeds differing in muscularity. *BMC Developmental Biology* **7**, 1-16.

Paixão D.M., Guimarães S.E.F., Silva Filho M. I., Lopes P.S, Pereira M.S., Sollero B.P. (2009) Detecção de locos de características quantitativas nos cromossomos 16, 17 e 18 de suínos. *Brazilian Journal of Animal Science* **37**, 1781-87.

Peixoto J.O., Guimarães S.E.F., Lopes P.S., Soares M.A.M., Pires A.V., Silva M.V., Almeida Torres R. & Almeida Silva M. (2006) Associations of leptin gene polymorphisms with production traits in pigs. *Journal of Animal Breeding and Genetics* **123**, 378-83.

Raymond F., Métairon S., Kussmann M., Colomer J., Nascimento A., Mormeneo E., García-Martínez C. & Gómez-Foix A.M. (2010) Comparative gene expression profiling between human cultured myotubes and skeletal muscle tissue. *BMC Genomics* **11**, 125.

SAS Institute (2007) *SAS OnlineDoc,Version 9.1.3.* SAS Institute Inc., Cary, NC, USA.

Shi L., Jones W.D., Jensen R.V., Harris S.C., Perkins R.G., Goodsaid F.M., Guo L., Croner L.J., Boysen C., Fang H., Qian F., Amur S., Bao W., Barbacioru C.C., Bertholet V., Cao X.M., Chu T.M., Collins P.J., Fan X.H., Frueh F.W., Fuscoe J.C., Guo X., Han J., Herman D., Hong H., Kawasaki E.S., Li Q.Z., Luo Y., Ma Y., Mei N., Peterson R.L., Puri R.K., Shippy R., Su Z., Sun Y.A., Sun H., Thorn B., Turpaz Y., Wang C., Wang S.J., Warrington J.A., Willey J.C., Wu J., Xie Q., Zhang L., Zhang L., Zhong S., Wolfinger R.D. & Tong W. (2008) The balance of reproducibility, sensitivity and specificity of lists of differentially expressed genes in microarray studies. *BMC Bioinformatics* **9**, S10.

Silva K. M., Paixão D.M., Silva P.V., Sollero B.P., Pereira M.S., Lopes P.S., Guimarães S.E.F. (2008) Mapping of quantitative trait loci and confirmation of the FAT1 region on chromosome 4 in an F2 population of pigs. *Genetics and Molecular Biology* **31**, 475-80.

Steibel J.P., Poletto R., Coussens P.M., Rosa G.J.M. (2009a) A powerful and flexible linear mixed model framework for the analysis of relative quantification RT-PCR data. *Genomics* **94**, 146-52.

Steibel J.P., Wysocki M., Lunney L.K., Ramos A.M., Hu Z.-L., Rothschild M.F. & Ernst C.W. (2009b) Assessment of the swine protein-annotated oligonucleotide microarray. *Animal Genetics* **40**, 883-93.

Sollero B.P., Paiva S.R., Faria D.A., Guimarães S.E.F., Castro S.T.R., Egito A.A., Albuquerque M.S.M., Piovezan U., Bertani G.R. & Mariante A. da S. (2009) Genetic diversity of Brazilian pig breeds evidenced by microsatellite markers. *Livestock Science*, **123**, 8-15.

Storey J.D. & Tibshirani R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440-45.

Sun L., Ma K., Wang H., Xiao F., Gao Y., Zhang W., Wang K., Gao X., Ip N. & Wu Z. (2007) JAK1-STAT1-STAT3, a key pathway promoting proliferation and preventing premature differentiation of myoblasts. *Journal of Cell Biology* **179**, 129-38.

Tang Z., Li Y., Wan P., Li X., Zhao S., Liu B., Fan B., Zhu M., Yu M. & Li K. (2007) LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. *Genome Biology* **8**, R115.

te Pas M.F.W., De Wit A.A.W., Priem J., Cagnazzo, M., Davioli R., Russo V. & Pool M.H. (2005) Transcriptome expression profiles in prenatal pigs in relation to myogenesis. *Journal of Muscle Research and Cell Motility* **26**, 157-65.

Tempelman R.J. (2005) Assessing statistical precision, power, and robustness of alternative experimental designs for two color microarray platforms based on mixed model effects models. *Veterinary Immunology and Immunopathology* **105**, 175-86.

Tintignac L.A., Lagirand J., Batonnet S., Sirri V., Leibovitch M.P. & Leibovitch S.A. (2005) Degradation of MyoD mediated by the SCF (MAFbx) ubiquitin ligase. *The Journal of Biological Chemistry* **280**, 2847-56.

Vianna A.T. Os suínos- Criação prática econômica. Série didáica no. 6. 2 ed. Serviço de Informação Agrícola, Rio de Janeiro. 1956.

Von der Hardt K., Kandler M.A., Fink L., Schoof E., Dötsch J., Bohle R.M., Rascher W. (2009) Laser-assisted microdissection and real-time PCR detect anti-inflammatory effect of perfluorocarbon. *American Journal of Physiology Lung Cell Molecular Physiology* **296**, 55-62.

Wang H.L., Zhu Z.M., Wang H., Yang S.L., Zhao S.H. & Li K. (2006) Molecular characterization and association analysis of porcine CA3. *Cytogenetic and Genomic Research* **115**, 129-33.

Wesolowski S.R., Raney N.E. & Ernst C.W. (2004) Developmental changes in the fetal pig transcriptome. *Physiological Genomics* **16**, 268-74.

Wigmore P.M. & Evans D.J. (2002) Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis. *International Review of Cytology* **216**, 175-232.

Wigmore P.M.C. & Stickland N.C. (1983) Muscle development in large and small pig fetuses. *Journal of Anatomy* **137**, 235-45.

Wilson M.E., Sonstegard T.S., Smith T.P.L., Fahrenkrug S.C. & Ford S.P. (2000) Differential expression during elongation in the preimplantation pig embryo. *Genesis* **26**, 9-14.

Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J. & Speed T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* **30**, 1-15.

Yelich J.V., Pomp D., & Geisert R.D. (1997) Detection of transcripts for retinoic acid receptors, retinal-binding protein and transforming growth factors during rapid trophoblastic elongation in the porcine conceptus. *Biology of Reproduction* **57,** 286-94.

Zhao S-H., Nettleton D., Liu W., Fitzsimmons C., Ernst C.W., Raney N.E. & Tuggle C. K. (2003) Complementary DNA macroarray analyses of differential gene expression in porcine fetal and postnatal muscle. *Journal of Animal Science* **81**, 2179-88.

CHAPTER II
(ARTICLE II)

Expression patterns evaluation of functional genes during skeletal muscle development in two genetic groups of pigs differing in muscularity

B.P. Sollero[†*], F.F. e Silva[¥], C.S. do Nascimento[†], D.M. Paixão[†], C.W. Ernst[*], P.S. Lopes[†] and S.E.F. Guimarães [†]

[†]Animal Science Department, Universidade Federal de Viçosa, Viçosa, 36571-000, Brazil; [¥]Statistics Department, Veterinary Department, Universidade Federal de Viçosa, Viçosa, 36571-000, Brazil,

[*]Department of Animal Science, Michigan State University, East Lansing, 48824, USA

Abstract

The interest in muscle growth potential is generally associated with characteristics determined prenatally during myogenesis, when the muscle fibers number is predominantly defined. The understanding of gene expression changes during prenatal pig muscle development is still limited. In this study fourteen genes identified as differentially expressed between breeds and/or developmental stages in a previous microarray research were chosen according to its biological functions during myogenesis process. The expression profile of these genes during prenatal (21, 40, 70 and 90 dpc) and early postnatal (10 wks) stages of the *longissimus dorsi* muscle development was observed between a Brazilian local pig breed (Piau) and a commercial line (Landrace x Large-White), two genetic groups phenotipicaly distincts in terms of lean muscle and muscularity. Based on qRT-PCR analyses using SYBER green method, a new reference gene was proposed to normalize this type of comparison in pigs (DDIT3). Nine genes were differentially expressed between both genetic groups ($P < 0.05$) and significant different levels of expression were also observed through the five time points. The results suggested that most of the relevant variations in mRNA levels of the analyzed genes seem to follow temporal waves of gene expression. Significant changes of transcription between the two groups were observed specially at 40 and 70 days, the main phases of the pig muscle development, but even higher significant changes between developmental stages were seen from 21d to 40d and from 90d to 10wk, especially. Expression patterns of genes related to cell proliferation, cell differentiation, energy metabolism and maintenance were distincts in a breed-specific way. Moreover speculative results indicated interdependencies between myogenesis and adipogenesis processes on the muscle type determination of fatty or lean muscle pig breeds. These results provide new data on developmental changes of expression profile of fourteen genes involved in different functional pathways related to prenatal stages and an early postnatal period of the myogenic processes in Piau and commercial pigs.

Key words: qRT-PCR, myogenesis, Piau.

Resumo

O interesse pelo potencial de crescimento de determinado músculo é geralmente associado a características determinadas durante o período pré-natal, denominado miogênese, quando o número de fibras musculares é predominantemente definido. O entendimento de mudanças na expressão gênica durante o desenvolvimento pré-natal do músculo de suínos ainda é limitado. Neste estudo, quatorze genes identificados como diferencialmente expressos entre raças e, ou estádios de desenvolvimento numa prévia pesquisa com microarranjos, foram escolhidos de acordo com suas funcionalidades envolvidas com processos biológicos decorrentes durante o processo de miogênese. O perfil de expressão destes genes durante períodos pré-natais (21, 40, 70 e 90 dias) e pós-natal (10 semanas pós-nascimento) do desenvolvimento do músculo *longissimus dorsi* foi observado entre uma raça brasileira local (Piau) e uma linhagem comercial (Landrace x Large-White), dois grupos genéticos de suínos fenotipicamente distintos em termos de capacidade de crescimento muscular e teor de carne magra. Baseado em análises de qRT-PCR pelo método SYBR green, um novo gene referência (DDIT3) foi proposto para realizar este tipo de comparação em suínos. Nove genes mostraram-se diferentemente expressos entre os dois grupos genéticos ($P<0,05$) e significativos níveis diferenciais de expressão foram também observados por entre os cinco estádios de desenvolvimento. Os resultados sugerem que variações dos níveis de mRNA mais relevantes dos genes analisados parecem seguir ondas temporais de expressão gênica. Mudanças significativas de transcrição entre os dois grupos genéticos foram observadas especialmente aos 40 e 70 dias de gestação, as duas fases mais importantes do desenvolvimento muscular de suínos. Mudanças ainda mais significativas entre os estádios do desenvolvimento foram vistas entre 21d e 40d, bem como entre 90d e 10 semanas pós-natal, especialmente. Os padrões de expressão de genes relacionados à proliferação celular, diferenciação celular, energia de metabolismo e manutenção da formação muscular foram distintos de forma raça-específica. Além do mais, resultados especulativos indicam interdependência entre os processos de miogênese e adipogênese na determinação de raças com maior porcentagem de gordura ou com maior conteúdo de carne magra. Estes resultados provêm novos dados sobre mudanças do perfil de expressão de quatorze genes envolvidos em diferentes vias metabólicas relacionadas a estádios pré-natais e um período pós-natal inicial do processo de miogênese em suínos brasileiros comerciais e naturalizado Piau.
Palavras chave: qRT-PCR, miogênse, Piau.

Introduction

More recently, the pig's role has expanded beyond just being a food source to potentially serving as an important model system for human health (Rothschild et al., 2004). In this way, more detailed studies of the myogenesis process 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). The lean muscle growth potential of an animal largely depends on the number of muscle fibers prenatally formed, because the postnatal increase in muscle fiber size is limited by genetic and physiological reasons (Rehfeldt et al., 2004). One of the unique characteristics of the skeletal muscle is its diversity on its morphological and biochemical properties (Ryu et al., 2006). In addition, pig breeds or genetic groups differing in muscle traits such as muscularity, fiber types and overall fatness, consequently result in different patterns of cell proliferation, differentiation, energy metabolism and muscle structure formation during the developmental process.

Considerable increased in publications regarding evaluation of transcriptome changes between breeds differing in muscularity and growth rate have been seen (Zhao et al., 2003; Te Pas et al., 2005; Te Pas et al., 2007; Tang et al., 2007; Lin and Hsu, 2005; Muráni et al., 2007; Cagnazzo et al., 2006; Lobjois et al., 2008; Li et al., 2008) but the orchestra of multigenes co-expressed and co-repressed which takes place during myogenesis process resulting in distinct muscle phenotypes, is still not completely elucidated.

Comparative expression profiling by competitive hybridization of microarrays in our previous analysis revealed differential expression of genes between the Piau local Brazilian pig breed and the white YorkshirexLandrace crossbred in muscle tissue at two prenatal ages (40 and 70 days pos conception) (Sollero et al., 2010-accepted). However, many genes activated in earlier stages of myogenesis in pigs are intrinsic for the continuation of this process (Te pas et al., 2005). In addition, when the fiber formation ceases, which is around 85 to 90 d (Wigmore and Stickland, 1983), the total number of fibers is theoretically established and the process of maturation of the myotubes into myofibers starts, finishing up in the early postnatal periods (Ashmore et al., 1973; Swatland and Cassens, 1973).

Interested so, in to expand the knowledge of muscle transcriptome between distinct types of breeds or genetic groups, this study used part of later information to continue investigate myogenesis process in porcine species. The Piau local pig breed as well a white composite pig line herded at the Breeding Farm of the *Universidade Federal de Viçosa*, differ considerably in growth rate, body composition, and muscle mass content. The first is more "obese" and the second more "lean". Instead of focusing in only two time-points of this process (40 and 70 dpc- days pos conception) - theoretically targeting both fiber differentiation waves - an evaluation of this dynamic biological process covering four prenatal ages and one early postnatal period is proposed in the present study. For that, fourteen genes potentially influencing myogenesis regulation and already found as differentially expressed in previous microarray study (Sollero et al., 2010-accepted) were selected. Differential expression profile evaluation based on quantitative Real Time PCR is proposed to investigate in more details the mechanism of muscle development and infer about up and down-regulations of particular genes related to cellular growth and proliferation, energy metabolism, tissue morphology, cellular differentiation and maintenance.

Material and methods

Animals

Three samples of RNA extracted from the *Longissimus dorsi* (LD) muscle of Piau and white composite fetuses represented by each gilt (n=3/ age) at 21, 40, 70 and 90 d of gestation and at one postnatal age (30kg at 10 weeks after birth) were used in the current qRT-PCR analysis. In the total, 30 samples represented the five time-points analyzed. Both genetic groups remained at the Pig Breeding Farm from the *Universidade Federal de Viçosa* (Brazil) and the sample procedure were done in accordance with the law for the protection of animals approved by the institutional animal welfare and ethic/protection commission of the Medicine Veterinary Department of the same University.

Muscle samples from the *Longissimus dorsi* of each fetus/piglets were collected with sterile blades and stored in RNAlater solution at 4℃ overnight. After that, samples were transferred to - 70℃ until RNA procedure and for undetermined conservation time.

Visible subcutaneous fat in Piau fetuses (from 40 d to 90 d of gestation) and piglets (30 kg) was carefully removed from the muscle samples.

RNA extraction and cDNA synthesis

Total RNA from approximately 30 mg RNAlater stabilized LD tissue from the Piau and the white composite fetuses/piglets was isolated using the RNeasy Mini kit with on-column RNase-free DNase digestion (Qiagen) according to the manufacturer's protocol. RNA concentrations were checked by NanoVue Plus Spectrophotometer (GE Healthcare) with an optimal 260/280 ratio between 1.8 and 2.1. Purity and integrity were determined with an Agilent RNA 6000 Nano Kit using the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.). Total RNA (1 µg) was reverse transcribed with oligo(dT) primers using the ProtoScript® M-MuLV First Strand cDNA Synthesis Kit (New England, BioLabs®$_{Inc}$). All the reaction conditions were performed as recommended by the manufacturer.

Quantitative real-time PCR

Fourteen genes were selected beforehand based on their biological functions suggested by networks generated with genes differentially expresses corresponding to contrasts (between ages and/or breeds) presented in previous microarray experiment (Sollero et al., 2010-accepted). These fourteen genes were separated in 3 functional groups according to Gene Ontology: signal transduction/transcription factor activity and/or growth factor activity (*CTNNB1* -β-catenin; *GRN* –Granulin; *HOXA5* -homeobox A5, *HSP90B1* -heat shock protein 90kDa beta (Grp94), member 1 and *MSTN* -Myostatin), reporter signaling pathway and/or metabolic process (*PPARGC1A* -peroxisome proliferator-activated receptor gamma, coactivator 1; *FABP4* -Fatty acid-binding protein 4 and *GAPDH* -Glyceraldehyde 3-phosphate dehydrogenase) and regulation of muscle contraction and/or fiber formation (*CAV3* -Caveolin-3*; CSDA* -Cold shock protein A*; DLK1* -Delta-like 1 homolog- *Drosophila; TNC* –Tenascin*; PAK1* -p21 protein (Cdc42/Rac)-activated kinase 1 and *TRIM63* -tripartite motif-containing 63). The primers were designed using PrimerQuest software (www.idtdna.com/Scitools/Aplications/PrimerQuest) from IDT Scitools

(Integrated DNA Technologies, Inc. Iowa City, IO, USA). When mRNA sequences for studied genes from pigs (*Sus scrofa*) were unavailable, sequences from humans (*Homo sapiens*) were used instead (ST 1). Interestingly, the gene DDIT3 (DNA-damage-inducible transcript 3) firstly selected to be evaluated through the five developmental stages, once it was found as differentially expressed between breeds at 70 days of gestation in previous analysis (Sollero et al., 2010-accepted), showed to be constitutively expressed between both current genetic groups evaluated and through the prenatal and postnatal stages. So forth, this gene was properly included as a reference gene in the normalization analyses.

Quantitative real-time PCR reactions were performed with SYBR® Green PCR Master Mix (Applied Biosystems) on an ABI Prism 7300 sequence detection system (Applied Biosystems). The first step of the reaction comprehended of 95℃ for 10 minutes, following the second step setting up for 40 cycles at the same temperature for 15 seconds and a final extension at 60℃ for one minute. The efficiency of each gene was accessed in order to choose the best combination of cDNA and primer concentration in the following reactions. Once the slope of a linear regression (established between the *log* cDNA and the C_q values) for each primer concentration was obtained, the PCR efficiency (E) was calculated from the following formula: E = 10exp(-1/slope). The slope values of the genes (the fourteen targets and the reference one) ranged from -2.70 to -3.80 corresponding to values of amplification efficiency from 0.8 to 1.32 (Sup. table 1 – Appendix 2). All the reactions were done in duplicates in a reaction of final volume of 20µ*L*.

Statistical analysis

In the current experiment, the Piau breed and the white composite pig line represented the two levels of the first factor (genetic groups) and the four time-points of prenatal ages (21, 40, 70 and 90 days post conception) plus the postnatal age at 30kg live weight, represented the five levels of the second factor (age). Thirty samples representing both factors were analyzed for each fourteen target genes and for the DDIT3 reference gene. In order to perform a statistical method that allows accommodation of more complex experimental designs and testing general hypotheses assuming independent random effect for the

control and target genes in each biological replicate, the %QPCR_MIXED macro developed in SAS (v. 9.1.3) was applied (Steibel at al., 2009). Based on a linear mixed model this software analyzed the quantification cycle (C_q) data of each sample in each gene of interest as a completely randomized design with two factors.

$$y_{gikr} = TG_{gi} + B_{gik} + D_{ik} + e_{gikr}$$

Where y_{gijkr} is the C_q obtained from the thermocycler for the *gth* gene (target or DDTI3) from the *rth* well corresponding to the *kth* animal subjected to the *ith* treatment; TG_{gi} is the effect of treatment *i* in the expression of gene *g*, B_{gik}~N(0, σ^2_{Bg}) is the gene-specific random effect of the *kth* animal, D_{ik}~N(0,σ^2_D) is the random sample-specific effect (common to both genes) and e_{gikr}~ N(0, σ^2_e) is the residual term. The treatments (*i*) consisted of the combination of two factors: genetic group and developmental stage (age).

Fold change values (relative expression measures) were obtained by transforming the estimates (C_q) detected after statistical analyses in that program: Fold change=$2^{-Estimate}$. The "estimates" generated as results from the statistical analyses are the differences obtained in C_q values between the two levels under evaluation (contrast). Once the efficiency (*E*) of the qRT-PCR reaction was close or equal to 2.0, one cycle (C_q) of difference between two levels (or samples) means 2.0 fold change of difference, or still, twice as much expression in the first level in comparison with the second level. Also, understanding that higher C_q values means lower transcriptional expression of a specific gene for a specific sample relatively to others and that lower C_q values means higher transcriptional expression, negative values of "estimates" indicate a positive fold change (relative expression) of the first level relatively to the second under investigation, after applying the transformation (Fold change=$2^{-Estimate}$).

Results

The results obtained by C_q (quantification cycle) values of each target gene in each genetic group and developmental stage before statistical analyses (not normalized) are shown by graphs in the supplementary figure 1 (Appendix 2). These graphs help to understand the sense of gene regulation and especially the differences in expression between both genetic groups through the five time-points. Similarly to the first supplementary figure, the figure 1 represents the fold changes of the statistically significant estimates results comparing the white composite against the Piau breed observed in each developmental stage for each target gene. The figure 2 also represents the fold change results but now comparing developmental stages two-by-two within each genetic group (columns represent the relative expressions in each genetic group).

The entire statistical results of each contrast between genetic groups within each developmental stage (n=5) and between development stages within each genetic group (n=8) can be checked in the supplementary table 2 (Appendix 2).

During the prenatal stage, the myostatin (MSTN) gene showed to be slightly more abundant at 40 d of gestation in both genetic groups, although the white composite pigs expressed 4.17 fold more than the Piau breed ($p<0.05$) at 21 d of gestation, 7 fold more ($p<0.05$) at 40 d and 6.49 fold more ($p<0.05$) at 70 d of gestation. We also observed significant decreased expression from 90 d of gestation to the early postnatal stage (at 30kg live weight) in both genetic groups ($p<0.001$).

The expression of the Hoxa5 gene was significantly down-regulated in the Piau breed from 21 d to 40 d of gestation ($p<0.5$), while more constantly expressed and slightly up-regulated until 40 d in the white breed. After 90 d of gestation, the expression decreased significantly in both genetic groups ($p<0.5$), but drastically in the Piau breed (28.97 fold change, $p=0.001$).

The β-catenin gene presented to be significantly down-regulated in both genetic groups during the prenatal stage and then up-regulated from 90 dpc to 30kg after birth. This gene was significantly more expressed in the white composite line at 70 d ($p<0.001$) and at 90 d of gestation ($p=0.08$) in comparison with the Piau breed.

In our study there was a trend of higher expression of the PPARGC1A (peroxisome proliferator-activated receptor gamma, coactivator 1) gene in the Piau breed at 21 d ($p<0.1$) but significantly more expressed at 70 d of gestation ($p<0.05$), related to the white composite line. A peak of expression was found later and in lower magnitude in the white composite line (90 d relatively to 70 d, FC=5.6503, $p<0.1$) in comparison with the Piau breed (70 d relatively to 40 d, FC=6.2803, $p<0.05$). At 30kg both genetic groups presented decreased expression ($p<0.05$).

The Fatty acid-binding protein 4 (FABP4) gene was up-regulated from 21 d of gestation to 30kg live weight in the Piau breed, presenting two significant peaks of expression at 40 d ($p<0.001$) and at 90 d ($p<0.05$) of gestation. Otherwise, in the white composite pigs, the peaks were earlier: at 40 d ($p<0.05$) and 70 d ($p<0.05$) of gestation. In this way, this gene was significantly more expressed in the white pigs at 70 d (3.95 fold change, $p<0.05$) in comparison with the Piau breed.

According to the results, it was clear that such gene like GAPHD could not be used to normalize the expression of the other genes analyzed in this type of study. This gene presented to be down-regulated until 70 d of gestation, and after that significantly up-regulated until 30 kg after birth in both genetic groups. There was a trend of expression in the white composite line to be higher at 70 d ($p<0.01$), but significantly higher at 90 d of gestation (FC= 9.5, $p=0.009$) in comparison with the Piau breed.

The TRIM63 gene (alias MuRF1- Muscle-specific RING finger protein 1) in the white pigs was significantly more expressed at 40 dpc (FC= 4.06; $p<0.05$), 90 dpc (FC= 7.13; $p<0.05$) and at the postnatal time-point (FC=3.93; $p<0.05$) in comparison with the Piau breed. Although in low level of expression in both genetic groups during the prenatal development, its expression was more expressive in the early postnatal period.

The expression pattern of the CAV3 gene (Caveolin-3) followed a parable shape in both genetic groups. A significant difference in expression between them was seen at 90 d of gestation, where there is a trend of peak of expression in both groups, but higher in the white composite line ($p=0.08$). The increased in expression from 21 d to 40 d of gestation was very prominent also in both groups ($p<0.001$).

The HSP90B1 gene did not show any significant difference of expression between both genetic groups. Its down-regulated expression pattern observed, was more significant (FC~10.0; $p<0.001$) from the end of the prenatal development (90 d) to the postnatal time-point (30kg) in both.

The Granulin gene (GRN) presented a similar pattern to the HSP90B1 gene. It was down-regulated expression pattern until 70 dpc, although slightly up-regulated from 70 to 90 dpc especially in the Piau breed ($p=0.003$). No significant difference was seen between both genetic groups.

Otherwise, the CSDA gene presented the opposite pattern during the prenatal development compared with these last two genes. It was up-regulated from 21 d to 90 d of gestation, where there was a peak of expression in the white composite line significantly higher ($p<0.001$) than in the Piau breed.

The PAK1 gene (p21 protein (Cdc42/Rac)-activated kinase 1) presented two peaks of expression in both genetic groups, firstly at 21 d and then at 70 d of gestation. Still, the white composite group was significantly more expressed than the Piau breed at 40 d ($p=0.057$). This gene had its expression decreased after 70 d of gestation in both genetic groups, especially from 90 dpc (days post conception) to 30kg ($p<0.001$).

The Piau breed presented a peak of expression of the TNC gene (tenascin C) at 40 dpc relative to 70 dpc ($p<0.001$). Even though presenting a similar expression pattern, the white composite group was significantly more expressed at 70 d ($p<0.00$) and 90 d of gestation ($p<0.05$) than the Piau. From 90 d to 30kg, there was a decreased expression in both genetic groups ($p<0.001$).

The DLK1 (delta-like 1 homolog (Drosophila)) gene presented the most interesting expression pattern among all the 14 genes evaluated. While almost 600 fold more expressed in the white pigs relatively to the Piau at 70 d of gestation ($p<0.001$), the beginning (21 d – 40 d) and the end (90 d – 30 kg) of the whole development stages evaluated here, were pretty much similar between both: firstly up-regulated (white composite line, $p<0.001$ and Piau $p<0.05$) and then down-regulated (both genetic groups with $p<0.001$).

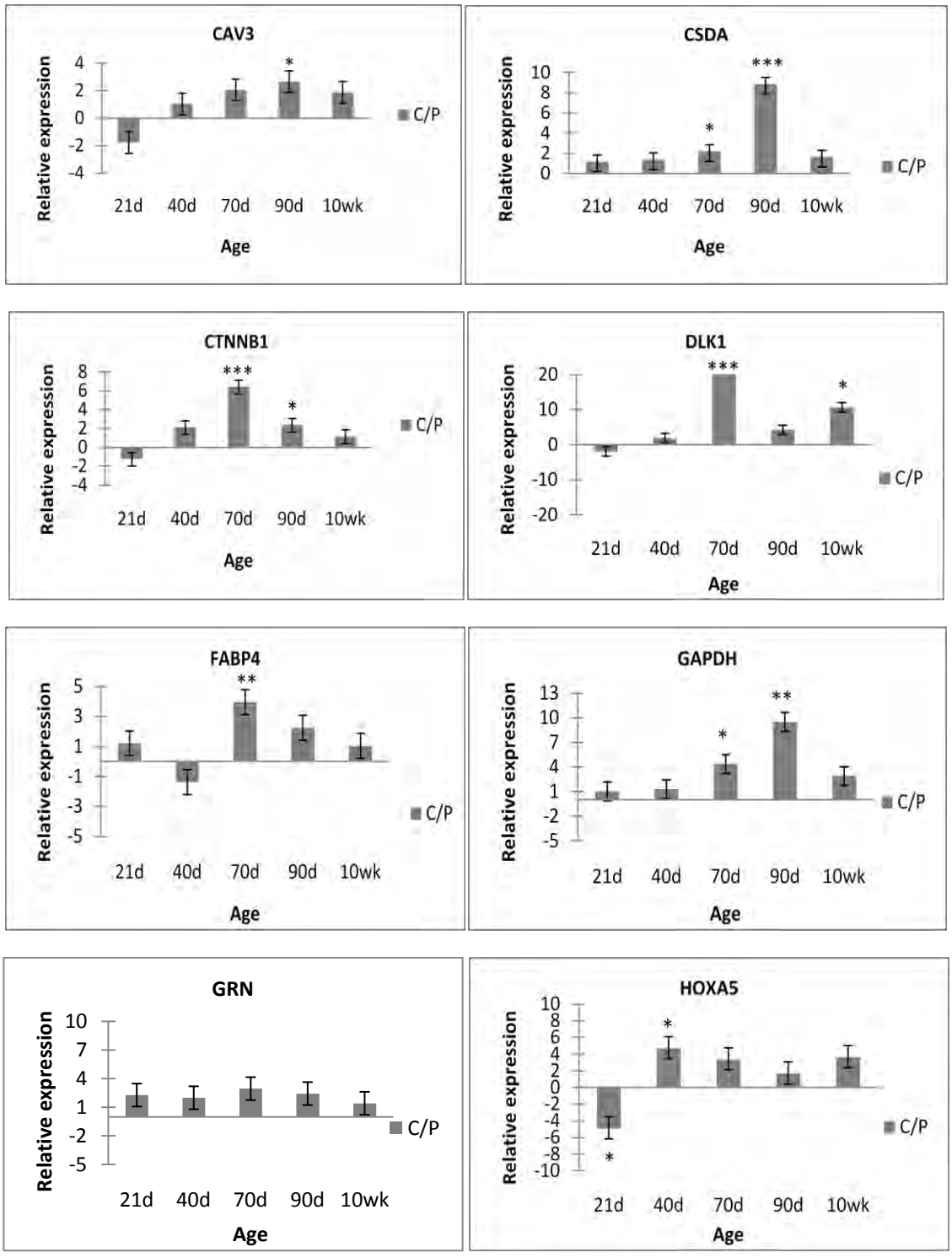


Figure 1- Relative expression (fold change) value between genetic groups (C/P=Composite line/ Piau) within each pre or postnatal age for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.

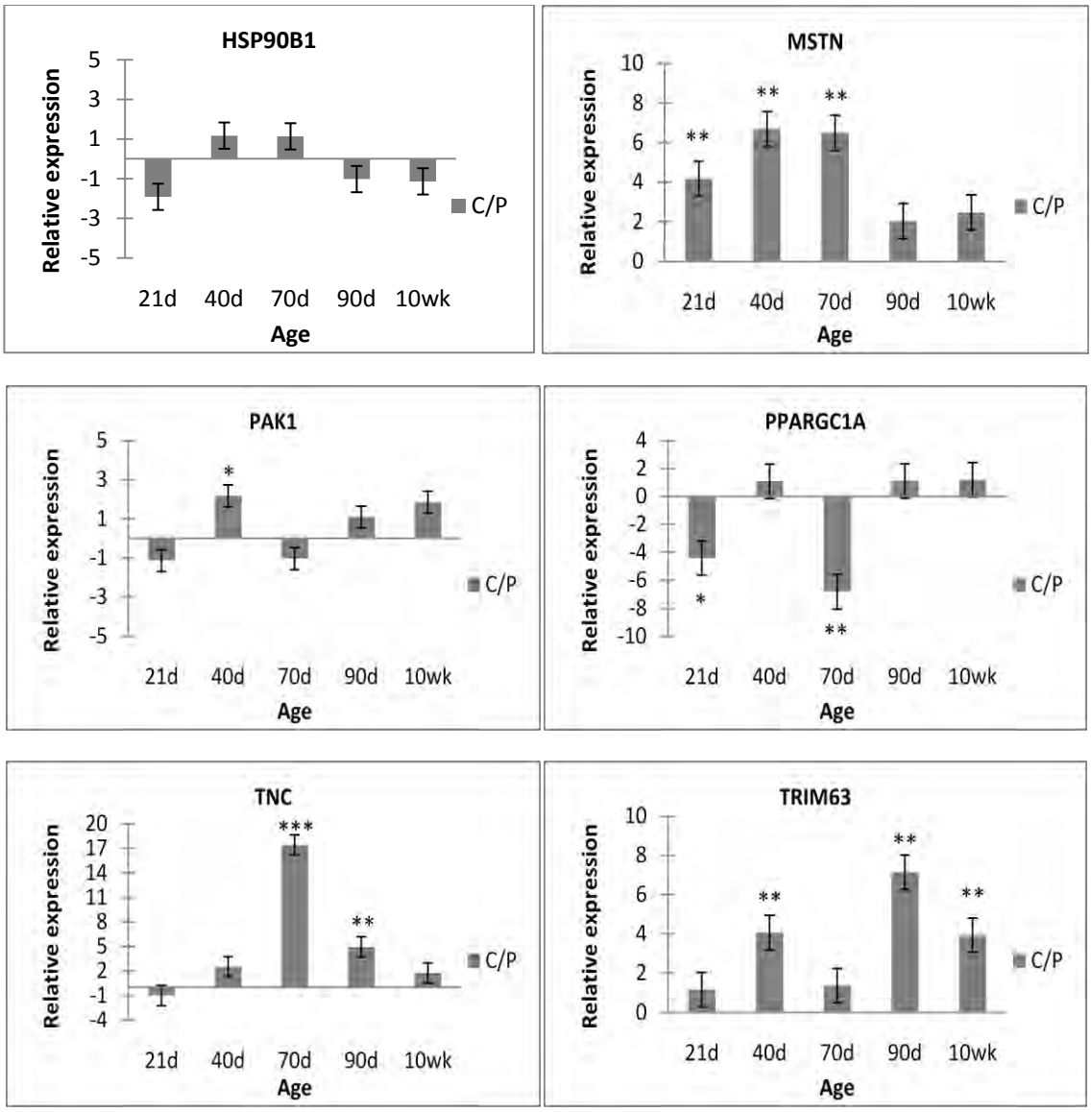


Figure 1 cont.- Relative expression (fold change) value between genetic groups (C/P=Composite line/ Piau) within each pre or postnatal age for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.

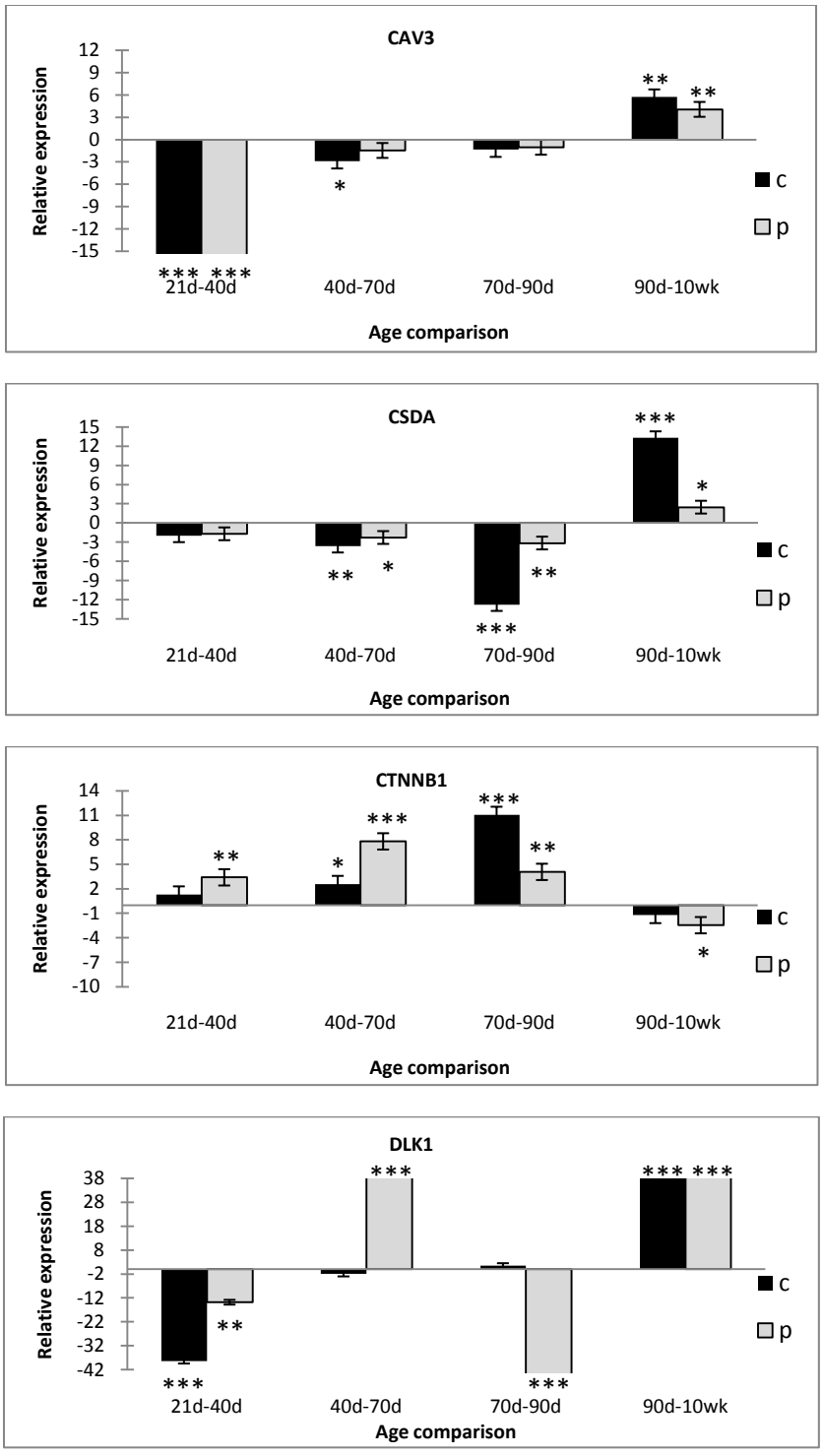


Figure 2- Relative expression (fold change) value between ages interval within genetic groups (c=Composite line; p=Piau) for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.

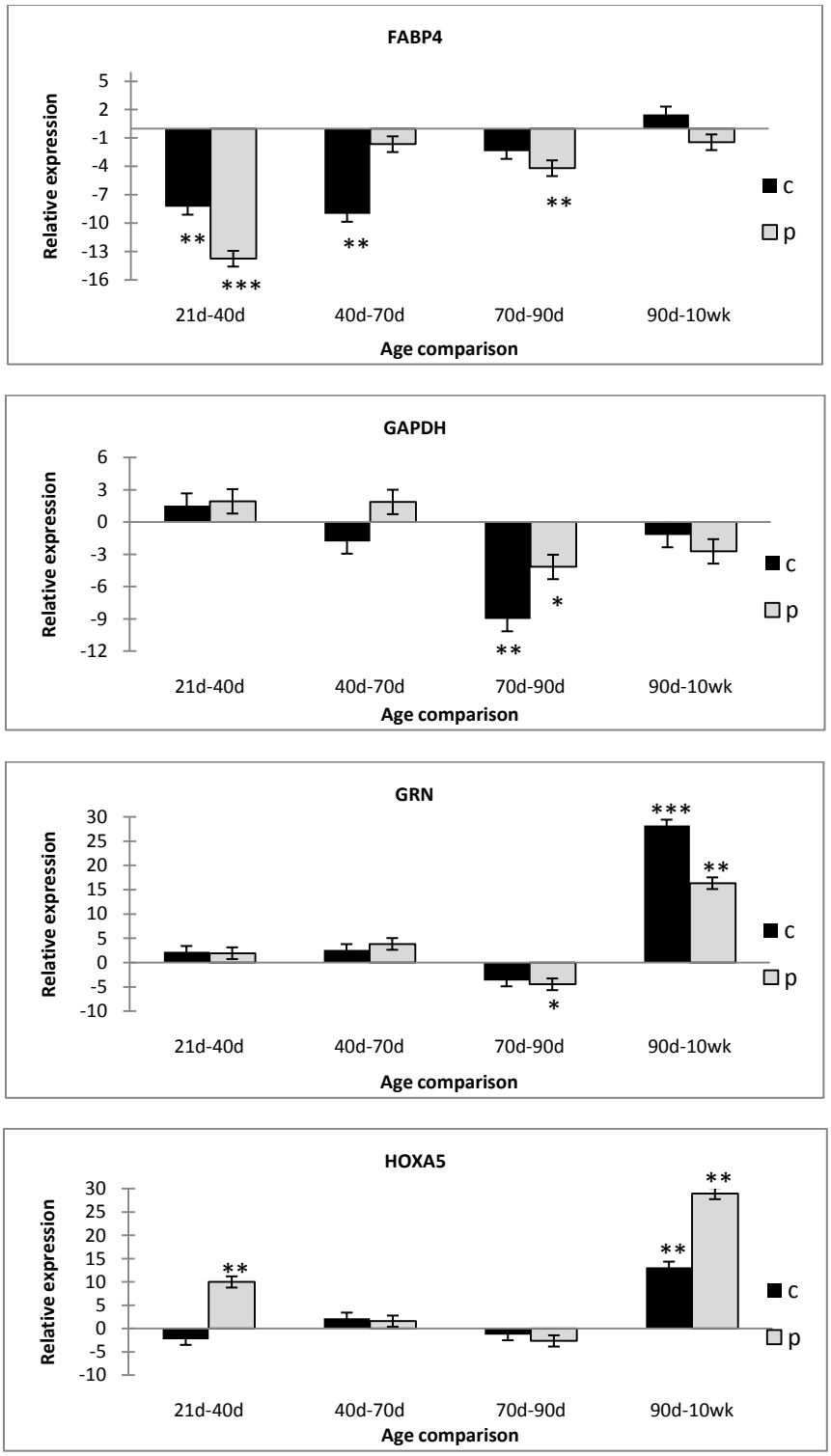


Figure 2 cont.- Relative expression (fold change) value between ages interval within genetic groups (c=Composite line; p=Piau) for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.

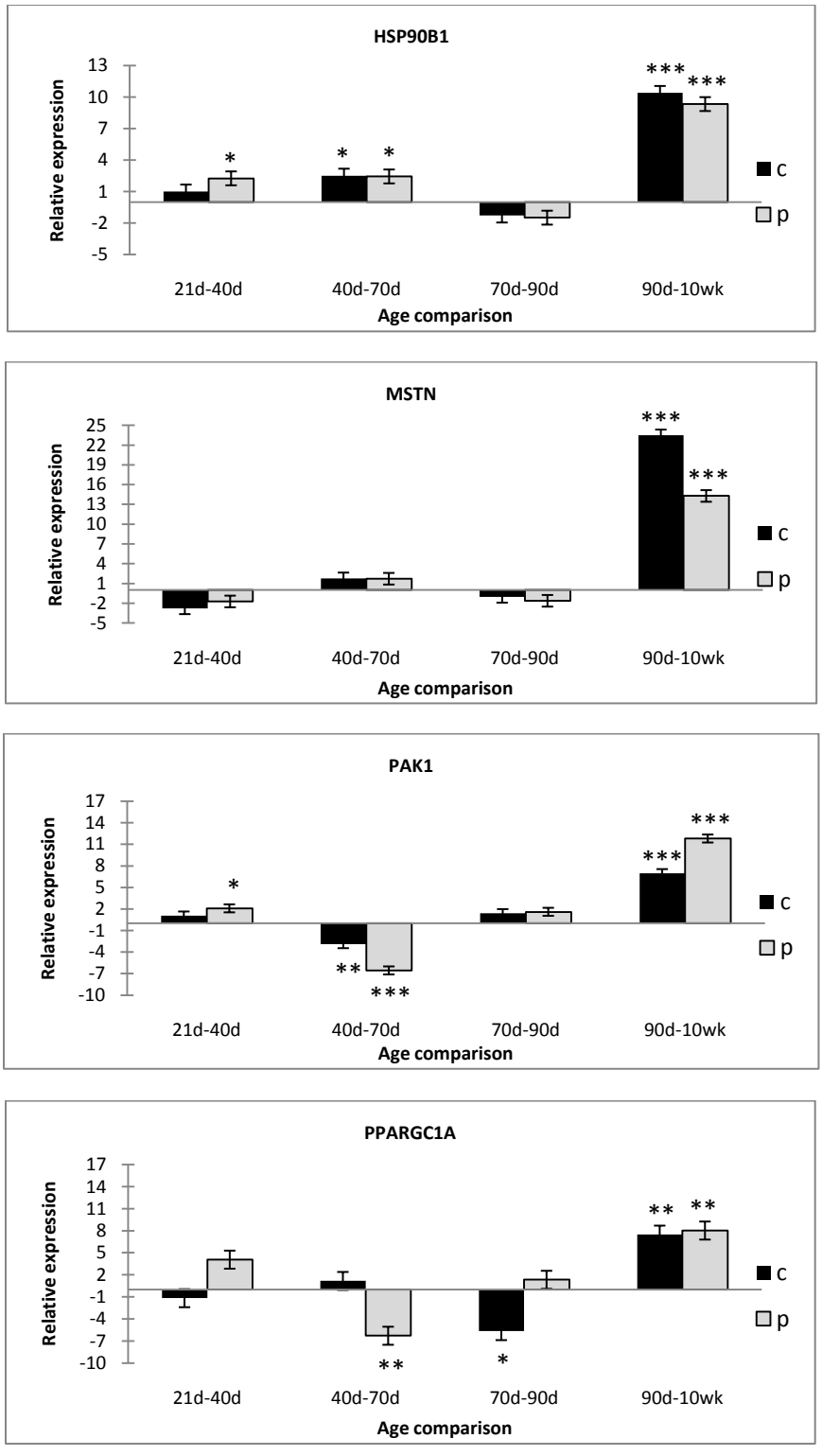
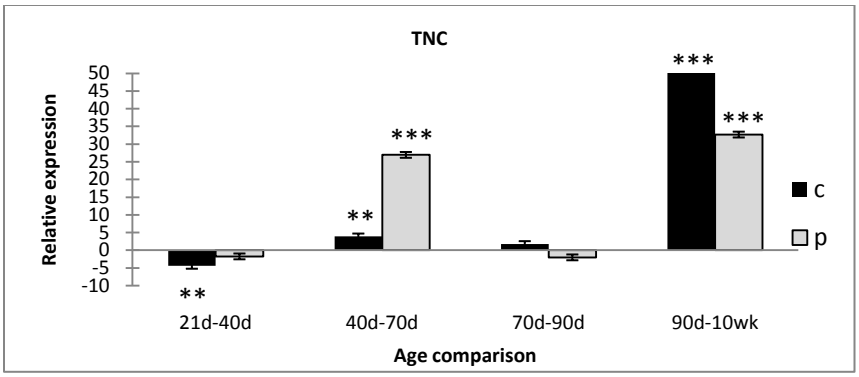


Figure 2 cont.- Relative expression (fold change) value between ages interval within genetic groups (c=Composite line; p=Piau) for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.



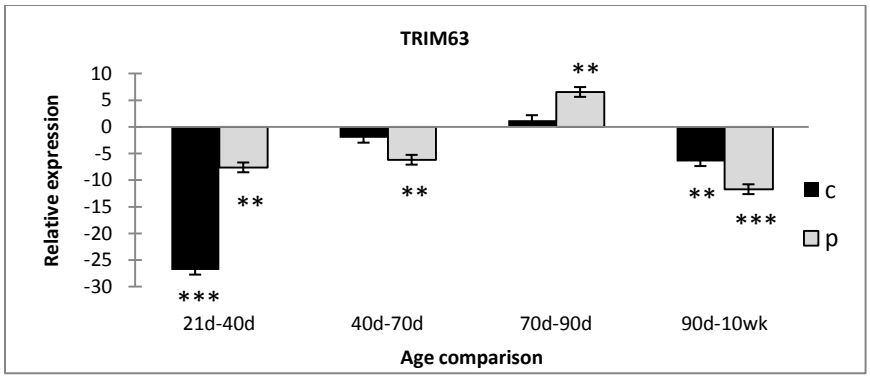


Figure 2 cont.- Relative expression (fold change) value between ages interval within genetic groups (c=Composite line; p=Piau) for each target gene. Columns without Standard Error bars, means FC value higher than 50.0. See sup. table 2. *p<0.1; **p<0.05; ***p<0.001.

Discussion

Computing valid P-values associated with any general linear hypothesis of interest is one of the main advantages of the method used to analyze this data, but specially the consideration of gene-specific biological effects by taking account the standard error (of C_q value in ABI) of each sample in each target and reference gene.

Cellular growth and Proliferation

Although extensively studied for its role as a negative regulator of skeletal muscle development (McPherron et al., 1997; Lin et al., 2002a; Grobet et al., 2003; Gilson et al., 2007), few reports describe the expression profile during prenatal development of the Myostatin gene in pigs (Ji et al., 1998; Patruno et al., 2008). This gene belongs to the transforming growth factors-β superfamily (TGF-β) (Joulia-Ekaza and Cabello, 2006), and it is not completely understood whether myostatin influences only myofiber formation or perhaps muscle growth and metabolism in association with conditions causing hypertrophy, arrested growth, or atrophy. Ji et al., (1998) found that myostatin mRNA abundance was low in *longissimus* muscle of pig fetuses at 21, 35 dpc and at birth, although markedly increased (2-fold, $p<0.05$) by 49 dpc. Moreover, the same authors found higher expression level of myostatin gene in *longissimus* muscle in lower-birth-weight piglets than in normal-weight piglets. Otherwise, this study showed almost the same magnitude of expression changes from 21 to 70 dpc in both genetic groups, although in higher expression, especially at 40 d of gestation, in the white composite line. More similarly to our results, Patruno et al., (2008) carried out an experiment on myogenic cells of Large White commercial pig breeds and found that the highest level of MSTN mRNA in the LD muscle was detected before birth (at 30 and 70 days of gestation), while the lowest level was observed in adults.

Theoretically, we would expect higher expression of this gene in the native breed due to its less final muscle mass content in comparison with the white composite pigs. Thus, how myostatin regulates myoblast proliferation, differentiation and/or fusion without negative affects the final muscle mass of

white composite pig lines, remains to be established, but during the period of myoblast proliferation Myostatin gene is more expressed in this genetic group. Similarly, but studying bovines, the myostatin RNA was expressed at significantly higher levels in the muscle of Piedmontese-sired fetuses (not manifesting the MSTN mutated gene) at 60 and 135 d of gestation (which represent primary and secondary myogenesis in this species, respectively) compared with the Wagyu-sired calves genetically predisposed to accumulate intramuscular fat (Lehnert et al., 2007).

The homeobox gene family specifies spatiotemporal gene expression patterns during development (Halayko and Solway, 2001; Tabaries et al., 2005). The Hoxa5 gene is a transcription factor with helix-turn-helix DNA binding motif that potentially interacts to antiapoptotic proteins. Like described by Stasinopoulos et al., (2005), Hoxa5 protein binds to the p53 promoter and activates p53 expression. Expression of the p53 gene protects cells against malignant transformation, but also can inhibits cell proliferation. Others results also suggested that putative TFBSs (Transcription Factors Binding Sites), such as Hoxa5 (homeobox A5), were identified in and adjacent to the SNP in the bovine Myf5 gene fragment, affecting Myf5 transcription in adult bovine muscle (Robakowska-Hyzorek et al., 2010). Transgenic mice that over expresses Hoxa5 gene shows 12-fold increase of the Igfbp1 gene and a 2-fold down-regulation of Igf1 gene expression in liver (Foucher et al. 2002), both genes being directly engaged in muscle development. In this way, extrapolating these results for the context of prenatal development, we suggest that the Hoxa5 gene may be indirectly involved in myoblasts proliferation - potentially protecting cells or regulating cell proliferation in earlier stages of the muscle development. Although suggested that there is no essential role for p53 during development of skeletal muscle *in vivo* (White et al., 2002), it is possible that activated by HOXA5, p53 may "interrupts" myoblast proliferation, accelerates differentiation and subsequently myotube formation prenatally; specially in the native breed, but this hypothesis still needs to be validated.

Reported as regulators of cellular proliferation that modulates epithelial cell function, the family of putative growth factors called granulins (Bateman et al., 1990) presented highly expressed at 21 d of

gestation in both genetic groups. According to the description by Gene Ontology, different members of the granulin protein family may act as inhibitors, stimulators, or have dual actions on cell growth.

Some results (Sato et al., 2000) indicated that HSP90 molecule binding to Akt (key signaling kinase (Akt1) in muscle development), protects cells from undergoing apoptosis by preserving phospho-Akt level and Akt kinase activity. Most of the identified cellular targets of HSP90 are signaling proteins. According to the same authors, proteins from the HSP90 gene family act as chaperone for unstable signal transducers and keeps them poised for activation until they are stabilized by conformational changes associated with signal transduction. Not differentially expressed between the two genetic groups evaluated here, this gene appears to be intrinsic during myogenesis process, especially in the beginning, due to the high expression level presented. Heat shock proteins, which are also indicators of cellular stress (Liu and Steinacker, 2001) were detected highly transcribed in Duroc in comparison with a native pig breed (Taoyuan) (Lin and Hsu, 2005). In bovines, the heat shock 70 kDa protein 1B (HSPA1B) and the heat shock 90 kDa protein 1, alpha (HSPCA) appeared up-regulated during the whole process of myogenesis until birth (Lehnert et al., 2007).

β-Catenin is an essential component of both intercellular junctions and of the canonical Wnt signaling, which plays critical roles in the development of multiple tissues through the regulation of cell proliferation, differentiation, and movement (Polakis, 2000; Yang et al., 2008). Through the Wnt/β-Catenin pathway, this gene acts as a transcriptional co-activator which is essential for myogenesis indirectly increasing the expression of MRFs, including Myf5 and MyoD (Cossu and Borello, 1999; Armstrong and Esser, 2005). Our results suggest that β-Catenin is more "expressive" during earlier periods of muscle development in both genetic groups, presenting even more active from 40 to 90 dpc in the composite line. Cagnazzo et al., (2006) found that the expression of the β-catenin in Pietrain fetuses was greater than in Duroc fetuses, except at 21 dpc, evaluating a proliferation-stimulating subgroup of genes involved in myogenesis. These results are similar to ours, although no statistical significance was seen for the higher expression in the Piau breed at 21 d of gestation.

Energy Metabolism and Tissue Morphology

The PPARGC1A gene, an interesting candidate gene for meat quality (Jacobs et al., 2006) which has many functions and can strongly differ between tissues (Erkens et al. 2008), is the master regulator of adipogenic differentiation (causes accumulation of intracellular lipids) (Spigelman et al., 2000) and plays a major role that regulates muscle fiber type determination (Lin et al. 2002b).

Still, some studies (Luo et al., 2009) also found that the Wnt/β-catenin signaling pathway alters adipocyte differentiation by inhibiting adipogenic genes expression, especially through the β-catenin gene. Even though manifested in earlier periods than 21 d of gestation, MSCs differentiate into mayoblast or adipocytes or fibroblasts (Du et al., 2010a) and it is interesting to relate the interaction between the PPARGC1A gene and the β-catenin gene (Yu et al., 2006) on this event. It is still not completely understood the action of PPARG in muscle compared with the adipose tissue to make conclusions (Guillerm-Regost et al., 2006), however, a potential relationship between these two genes (PPARGC1A and β-catenin) on fiber determination differentially between fatty and lean muscle pig breeds could also be suggested.

Overexpression and activation of PPARGC1A was previously shown to be followed by an increase of FABP4 expression and decrease levels of myogenic genes such as MyoD and Myogenin, promoting adipogenic conversion of myoblasts (Seale et al., 2008). Otherwise, Erkens et al., (2009) suggested a negative influence of PPARGC1A on FABP4 expression. The FABP4 gene is a candidate gene target downstream to PPARGC1A and higher expression levels were seen in FABP4 after both peaks of PPARGC1A in the Piau breed. The possibility of FABP4 acts on defining the early onset of intramuscular fat deposition in the native breed remains to be explored like discussed by Lehnert et al., (2007) studying myogenesis in bovine, as well, if FABP4 impairs some influence in differentiation of myoblasts, especially in white composite lines.

The glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene is involved in glycolysis process, and its upregulation after 70 dpc highlighted in the composite pig group, may emphasize the

energy uptake during muscle differentiation. Studying the regulation of glycogenolytic enzyme mRNA in porcine satellite cells, Henckel et al., (2007) observed that when satellites cells fuse into myotubes, the requirements for energy shifts from a survival to a activity-related type (which uptake huge amount of energy for muscle contraction) and that the level of GAPDH increases according to the activity of the glycolytic pathway, indicating the glycolytic capacity of muscle fibers of a certain muscle. This gene which participates in ATP metabolism indicates that at birth, the Piau and the composite line may differ less in energy metabolism. Otherwise, Cagnazzo et al. (2006) showed that the expression level of this gene was greatest in Pietrain fetuses, except for 21 d.

Oppositely as reported to Murani et al., (2007), but in agreement with Radonic´et al., (2004), the GAPDH gene could not be used as a reference gene in the present work due to its controlled expression pattern through the muscle development and between the two genetic groups.

Cellular differentiation and Maintenance

In vitro studies suggested that the expression of TRIM63 (tripartite motif-containing 63) is required for skeletal myoblast differentiation and myotube fusion, acting as a myogenic regulator of the microtubule network of striated muscle cells (Spencer et al., 2000). The TRIM motif is comprised of zinc-binding domains, a RING finger region, a B-box motif and a coiled-coil domain and is also known as MURF1 (Muscle-specific RING finger protein 1) (Zhang et al., 2006). The expression pattern seen in the white composite line suggests an active and more constantly involvement during the two most important stages of myogenesis for myoblast differentiation.and especially after birth (30kg), when maturation of muscle fiber is theoretically taking place. According to Handel, (1984) maturation of muscle fiber type by myosin ATPase staining was not seen until 4 wk postnatal.

Interestingly, MsCarthy et al. (2007) found that this gene was involved in some way with protein degradation circadian genes and so forth is of great interest given their central role in models of skeletal

muscle atrophy. The significant increased expression of TRIM63 seen from 40 to 70 dpc in the Piau breed corroborates with the proteolysis pathway discussed by Sollero et al. (2010-accepted).

The protein that contain "cold shock" domains (CSDs), like the cold shock domain protein A (CSDA), or MSY4 protein, was shown to be abundantly expressed during early- and mid-stage embryogenesis in mouse and also important for cellular proliferation and stress responses (Lu et al., 2006). In the present study, this gene was highly expressed during mid and final-stage of gestation in both genetic groups, suggesting to be more involved on maintenance and fiber repair after differentiation, especially in the white composite line evaluated. The protein coded by this gene was also shown to be related to that key signaling kinase (Akt1) in muscle development (Puente et al., 2006), already cited as related to HSP90B1 gene. Although not specified the function of this linkage (CSDA/Akt1), it was reported that stimulation of Akt1 by IGF-1 or insulin induces muscle hypertrophy, whereas inhibition of AKT1 by glucocorticoids promotes muscle atrophy.

Caveolin-3 expression is restricted to striated muscle tissue and during early muscle development is associated with the formation of T-tubules, suggesting involvement of caveolin-3 in the biogenesis of the T-tubule system during muscle differentiation of mouse skeletal muscle (Parton et al., 1997). Although pretty similarly regulated between both genetic groups of pigs here evaluated, CAV3 appears to be active during both myoblast differentiation waves and also in the early postnatal period. However, Parton et al. (1997) also inferred that caveolin-3 is not present in T-tubules in the final differentiated state. Also in mouse, an important role of myogenin in regulating caveolin-3 mRNA expression (positively) lately during differentiation was suggested (Biederer et al., 2000).

Expression of any form of Pak1 was shown to induce drastic changes in actin organization (Sells et al., 1999). These authors suggested that the p21-activated kinases (Paks) are important candidate effectors of reorganization of actin cytoskeleton and consequently structure mechanism of muscle contraction. Firstly activated at 21 dpc, this gene seems to be acting strictly during the beginning of both waves of fiber differentiation similarly in both genetic groups, although tending to be more expressed in the white composite line at 40 dpc.

Tenascin-C, an extracellular matrix component (ECM), modulates cell adhesion (Chiquet-Ehrismann et al., 1988). Fluck et al. (2000) demonstrated that upon loading of skeletal muscle, different spatial and temporal patterns of regulation are observed for ECM proteins such as tenascin-C. In fact, it is well known that tenascin-C mRNA and its protein levels are regulated by growth factors and hormones, like the TGF-β (Pearson et al., 1988).

The Delta-like 1 (DLK1) gene is well recognized as overexpressed in skeletal muscle of sheep with the callipyge phenotype (Charlier et al. 2001). Its role potentially during the second wave activating the myoblast differentiation in the composite line is highlighted in the present study. Otherwise, this gene also known as preadipocyte factor 1 (PREF-1), regulates cell fate determination and adipogenesis (Smas and Sul 1993). Samulin et al., (2008) observed that during porcine adipogenesis there was markedly decreased expression of two new variants DLK1C and DLK1C2 in visceral and subcutaneous adipose tissue. Investigating the LD skeletal muscle, the decreased expression at 70 d of gestation seen in the Piau breed, may suggest an opposed regulation of this gene between these two types of breed. Moreover, the expression pattern of the PPARGC1A, was also seen in opposite sense of the DLK1 gene. In this way, a hypothesis can be speculated: while myoblast differentiation is taking place, DLK1 is more expressed and early adipogenesis (prenatally) is potentially depleted. The role of DLK1 gene during development of lean muscle mass in pig was reported by Kim et al. (2004). According to these authors analyses of the parental origin of DLK1 polymorphism in the F2 offspring found that paternal inheritance of DLK1 allele 2 and maternal inheritance of the allele 1 was significantly associated with decreased fat deposition and increased lean muscle mass, while the opposite parental inheritance of these alleles was associated with slower prenatal and postnatal growth. Indeed, these results suggest that the polar overdominance mode of inheritance of the DLK1 gene (present in the pig chromosomal region homologous to the CLPG (Callipyge) locus in sheep) could be evaluated more precisely by distinguishing the parental origin of alleles at this locus.

Comparing with the previous microarray results (Sollero et al., 2010- accepted), the DLK1 gene also presented more highly expressed at 70 d of gestation in the Piau breed and in the crossbred relative to

40 d. However, no significant differential expression between both genetic groups was seen before. This is another indication that the mechanism in which this gene is acting during the muscle development in pigs is more complex and probably exhibits a breed/individual-specific expression pattern.

Others investigations about myogenesis process comparing genetic groups differing in muscle growth rate and capacity have suggested that fetuses from a more highly muscled breed of pigs show a later peak of myogenesis-related gene expression during fetal development (Te pas et al., 2005; Cagnazzo et al., 2006; Sollero et al., 2010-accepted). The present genes analyzed guide to take different conclusions, especially due to the potential mechanisms of activation and maintenance of myoblast differentiation stages during the muscle development of the white composite genetic group selected for lean muscle mass. Moreover, intrinsic mechanisms of myoblast proliferation differentially acting between these two groups of pigs still need to be understood.

Relating prenatal muscle development and intrinsic mechanisms of energy metabolism which interfere in the final muscle phenotype between different types of genetic groups, it is known that fat depot differences initiated during early embryogenesis has a "dominant effect" on the number of adipocytes existing inside the muscle - preferentially resulting in more intramuscular fat content (Du et al., 2010b). According to Vodovar et al. (1971) and Desnoyers and Vodovar (1974), the first elements of adipose tissue appear in the pig fetuses at about day 70 in periepididymal and pericardic depots, and at about day 106 in mesenteric depot. However, Piau fetuses from 40 d of gestation presented a well-developed subcutaneous fat layer similarly as reported by Cagnazzo et al. (2006) after observing Duroc pigs at 35 dpc. In this way, we instigate specific studies capable to conclude if there is some influence of adipose tissue formation during middle gestation on the myofibers formation of a muscle.

More specifically, to better understand the pivotal role of myostatin gene in skeletal muscle mass regulation during muscle development, an evaluation of fiber type composition in specific muscle tissue could be associated, like reported by Patruno et al. (2008). Moreover, we showed evidences that others myogenic regulators (more than the well known myogenic regulatory transcription factors (MRF) like

Myf5, MyoD and myogenin) of the microtubule network of striated muscle cells, which are linked with microtubule organization, most of the time, presented more activated in the composite line.

References

Ashmore, CR, PB Addis and L Doerr (1973) Development of muscle fibers in fetal pig. *Journal of Animal Science* **36**: 1088-1093.

Armstrong, DD and KA Esser (2005) Wnt/β-catenin signaling activates growth-control genes during overload-induced skeletal muscle hypertrophy. *American Journal of Physiology Cell Physiology* **289**, 853-59.

Bateman, A, D Belcourt, HPJ Bennett, C Lazure and S Solomon (1990) Granulins, a novel class of peptides from leukocytes. *Biochemical and Biophysical Research Communications*, **173**, 1161–1168.

Biederer, CH, SJ Ries, M Moseri, M Florio, MA Israel, F McCormick and R Buettneri (2000) The Basic Helix-Loop-Helix Transcription Factors Myogenin and Id2 Mediate Specific Induction of Caveolin-3 Gene Expression during Embryonic Development. The Journal of biological chemistry **275:**26245–26251.

Cagnazzo, M, MFW Te Pas, J Priem, AAC de Wit, MH Pool, R Davoli and V Russo (2006) Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. *Journal of Animal Science*, **84**, 1–10.

Charlier, C, K Segers, D Wagenaar, L Karim, S Berghmans S (2001) Human–ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. *Genome Research*, **11:** 850–862.

Chiquet-Ehrismann, R, P Kalla, CA Pearson, K Beck and M Chiquet (1988). Tenascin interferes with fibronectin action. *Cell* **53:** 383-390.

Cossu, G and U Borello (1999) Wnt signaling and the activation of myogenesis in mammals. *The EMBO Journal*, **18 (**24): 6867–6872.

Desnoyers, F and N Vodovar (1974) Apparition, origine et evolution des tissus adipeux epididymaire et pericardiaque du foetus de porc. *Annales de biologie animale, biochimie, biophysique* **14:** 769-780.

Du, M, X Yan, JF Tong, J Zhao and MJ Zhu (2010a) Maternal Obesity, Inflammation, and Fetal Skeletal Muscle Developmen. *Biology of Reproduction* **82**: 4–12.

Du, M, Y Jingdong, JZ Mei (2010b) Cellular signaling pathways regulating the initial stage of adipogenesis and marbling of skeletal muscle. *Meat Science* **86:** 103–109.

Erkens, T, K Bilek, A Van Zeveren and LJ Peelman (2008) Two new splice variants in porcine PPARGC1A. *BMC Research Notes* **1**:138. doi:10.1186/1756-0500-1-138.

Erkens, T, J Vandesompele, AV Zeveren, LJ Peelman (2009) Correlation between porcine PPARGC1A mRNA expression and its downstream target genes in backfat and longissimus dorsi muscle. *Journal of Applied Genetics* **50**: 361–369.

Flück, M, V Tunç-Civelek and M Chiquet (2000) Rapid and reciprocal regulation of tenascin-C and tenascin-Y expression by loading of skeletal muscle. *Journal of Cell Science* **113**: 3583-3591.

Foucher I, Volovitch M, Frain M, Kim JJ, Souberbielle JC, Gan L, Unterman TG, Prochiantz A, Trembleau A (2002) Hoxa5 overexpression correlates with IGFBP1 upregulation and postnatal dwarfism: evidence for an interaction between Hoxa5 and Forkhead box transcription factors. *Development* **129**:4065–4074.

Gilson, H, O Schakman, L Combaret, P Lause, L Grobet, D Attaix, JM Ketelslegers and JP Thissen (2007) Myostatin Gene Deletion Prevents Glucocorticoid- Induced Muscle Atrophy. *Endocrinology* **148**(1):452–460.

Guillerm-Regost, C, I Louveau, SP Se´bert, M Damon, MM Champ and F Gondret (2006) Cellular and Biochemical Features of Skeletal Muscle in Obese Yucatan Minipigs. *Obesity* **14:**1700- 1707**.**

Grigoriadis, AE, JN Heersche and JE Aubin (1988) Differentiation of muscle, fat, cartilage, and bone from progenitor cells present in a bone-derived clonal. *The Journal of Cell Biology* **106**:2139–2151.

Grobet, L, D Pirottin, F Farnir, D Poncelet, LJ Royo, B Brouwers, E Christians, D Desmecht, F Coignoul, R Kahn and M Georges (2003) Modulating Skeletal Muscle Mass by Postnatal, Muscle-Specific Inactivation of the Myostatin Gene. *Genesis* **35**:227–238.

Halayko, AJ and J Solway (2001) Plasticity in Skeletal, Cardiac, and Smooth Muscle Invited Review: Molecular mechanisms of phenotypic plasticity in smooth muscle cells. *Journal of Applied Physiology* **90**: 358–368.

Handel, SE (1984) Effects of low birth weight on postnatal development of skeletal muscle in the pig. Ph.D. Dissertation. University of Edinburgh, U.K.

Henckel, PR, PK Theil, IL Sorensen and N Oksbjerg (2007) Temporal changes in glycogenolytic enzyme mRNA during myogenesis of primary porcine satellite cells. *Meat Science* **75**: 248-255.

Jacobs, K, G Rohrer, M Van Poucke, F Piumi, M Yerle, H Barthenschlager, M Mattheeuws, A Van Zeveren and LJ Peelman (2006) Porcine PPARGC1A (peroxisome proliferative activated receptor gamma coactivator 1A): coding sequence, genomic organization, polymorphisms and mapping. Cytogenetics Genome Research **112**:106-113.

Ji, S, RL Losinski, SG Cornelius, GR Frank, GM Willis, DE Gerrard, FF Depreux and ME Spurlock (1998) Myostatin expression in porcine tissues: Tissue specificity and developmental and postnatal regulation. *American Journal of Physiology* **275**:1265– 1273.

Joulia-Ekaza, D and G Cabello (2006) Myostatin regulation of muscle development: molecular basis, natural mutations, physiopathological aspects. *Experimental Cell Research* **312**:2401–2414.

Kim K-S, J-J Kim, JCM Dekkers, MF Rothschild (2004) Polar overdominant inheritance of a DLK1 polymorphism is associated with growth and fatness in pigs. *Mammalian Genome* **15**: 552–559.

Lehnert, SA, A Reverter, KA Byrne, Y Wang, GS Nattrass, NJ Hudson and PL Greenwood (2007) Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Developmental Biology* **7**:95. doi:10.1186/1471-213X-7-95.

Li, M, X Li, L Zhu, X Teng, H Xiao, S Shuai, L Chen, Q Li and Y Guo (2008) Differential expression analysis and regulatory network reconstruction for genes associated with muscle growth and adipose deposition in obese and lean pigs. *Natural Science* **18**, 387-399.

Lin, CS, YC Wu, Y L Sun and MC Huang (2002a) Postnatal expression of growth/differentiation factor-8 (GDF-8) gene in European and Asian pigs. *Asian - Australasian Journal of Animal Sciences* **15**:1628–1633.

Lin, J, H Wu, PT Tarr, C-Y Zhang, Z Wu, O Boss, LF Michael, P Puigserver, E Isotani, EN Olson, BB Lowell, R Bassel-Duby and BM Spigelman (2002b) Transcriptional co-activator PCG-1α drives the formation of slow-twitch muscle fibers. *Nature* **418**: 797-801.

Lin, CS and CW Hsu (2005) Differentially transcribed genes in skeletal muscle of Duroc and Taoyuan pigs. *Journal of Animal Science* **83**: 2075-86.

Liu, Y and JM Steinacker (2001) Changes in skeletal muscle heat shock proteins: Pathological significance. *Frontiers in Bioscience* **6**:12–25.

Lobjois, V, L Liaubet, M SanCristobal, J Glenisson, K Feve, J Rallieres, P Le Roy, D Milan, P Cherel and F Hatey (2008) A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *International Society for Animal Genetics, Animal Genetics* **39**: 147–162.

Lu, ZH, JT Books, TJ Ley (2006) Cold shock domain family members YB-1 and MSY4 share essential functions during murine embryogenesis. *Molecular Cell Biology* **26**: 8410-8417.

Luo, X, H-X Li, R-X Liu, Z-S Wu, Y-J Yang and G-S Yang (2009) β-catenin protein utilized by Tumour necrosis factor-α in porcine preadipocytes to suppress differentiation. *BMB reports* **42** (6): 338-343.

McCarthy, JJ, JL Andrews, EL McDearmon, KS Campbell, BK Barber, BH Miller, JR Walker, JB Hogenesch, JS Takahashi and KA Esser (2007) Identification of the circadian transcriptome in adult mouse skeletal muscle. *Physioloy Genomics* **31**: 86–95.

McPherron, AC, AM Lawler, SJ Lee (1997) Regulation of skeletal muscle mass in mice by a new TGF-_ superfamily member. *Nature* **387**:83–90.

Muráni, E, M Murániová, S Ponsuksili, K Schellander and K Wimmers (2007) Identification of genes differentially expressed during prenatal development of skeletal muscle in two pig breeds differing in muscularity. *BMC Developmental Biology* **7:** 1-16.

Parton, RG, M Way, N Zorzi and E Stang (1997) Caveolin-3 Associates with Developing T-tubules during Muscle Differentiation. *The Journal of Cell Biology* **136**:137–154.

Patruno, M, F Caliaro, L Maccatrozzo, R Sacchetto, T Martinello, L Toniolo, C Reggiani and F Mascarello (2008) Differentiation Myostatin shows a specific expression pattern in pig skeletal and extraocular muscles during pre- and post-natal growth. *Differentiation* **76**:168–181.

Polakis, P (2000) Wnt signaling and cancer. *Genes & Development* **14**:1837–51.

Puente, LG, S Voisin, REC Lee and LA Megeney (2006) Reconstructing the Regulatory Kinase Pathways of Myogenesis from Phosphopeptide Data. *Molecular & Cellular Proteomics* **5**: 2244-2251.

Pearson, CA, D Pearson, S Shibahara, J Hofsteenge and Chiquet-Ehrismann R. (1988). Tenascin: cDNA cloning and induction by TGFbeta. *EMBO Journal* **7**: 2977-2982.

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

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

Rehfeldt, C, I Fieldler and NC Stickland (2004) 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*, p. 15 – 52. CABI Publishing: Oxon.

Robakowska-Hyzorek, D, J Oprzadek, B Zelazowska, R Olbromski, L Zwierzchowski. (2010) Effect of the g.–723G?T Polymorphism in the Bovine Myogenic Factor 5 (Myf5) Gene Promoter Region on Gene Transcript Level in the Longissimus Dorsi Muscle and on Meat Traits of Polish Holstein-Friesian Cattle. *Biochem Genetics* **48**:450–464.

Rothschild, MF (2004) Porcine genomics delivers new tools and results: This little piggy did more than just go to market. *Genetics Research, Cambridge* **83**, doi: 10.1017/S0016672303006621.

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

SAS Institute (2007) SAS *OnlineDoc*,*Version* 9.1.3. SAS Institute Inc., Cary, NC, USA.

Samulin J, PR Berg, H Sundvold, E Grindflek and S Lien (2008) Expression of DLK1 splice variants during porcine adipocyte development in vitro and in vivo. *Animal Genetics* **40:** 239–241.

Sato, S, N Fujita and T Tsuruo (2000) Modulation of Akt kinase activity by binding to Hsp90. *PNAS* **97**: 10832–10837.

Seale, P, B Bjork, W Yang, S Kajimura, S Chin, S Kuang, A Scimé, S Devarakonda, HM Conroe, HE Bromage, P Tempst, MA Rudnicki, DR Beier and BM Spiegelman (2008) PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454**: 961-967.

Sells, MS, JT Boyd and J Chernoff (1999) p21-Activated Kinase 1 (Pak1) Regulates Cell Motility in Mammalian Fibroblasts. *The Journal of Cell Biology* **145**:837-849.

Smas CM, HS Sul (1993) Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell* **73:** 725–734.

Sollero, BP, SEF Guimarães, VD Rilington, RJ Tempelman, NE Raney, JP Steibel, JD Guimarães, PS Lopes, MS Lopes and CW Ernst. (2010) Transcriptional profiling during fetal skeletal muscle development of Piau and Yorkshire-Landrace crossbred pigs. *Animal Genetics* (ID AnGen-10-05-0192).

Spencer, JA, S Eliazer, RL Ilaria Jr., JA Richardson and EN Olson (2000) Regulation of Microtubule Dynamics and Myogenic Differentiation by MURF, a Striated Muscle RING-Finger Protein. *The Journal of Cell Biology* **150:** 771–784.

Spiegelman, BM, Puigserver P, Wu Z (2000) Regulation of adipogenesis and energy balance by PPARϒ and PGC-1. *International Journal of Obesity* **24**:8–10.

Stasinopoulos, IA, Y Mironchik, A Raman, F Wildes, P Winnard Jr. and V Raman (2005) HOXA5-Twist Interaction Alters p53 Homeostasis in Breast Cancer Cells. *The Journal of Biological Chemistry* **280**, 2294–2299.

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

Swatland, HJ and RJ Cassens (1973) Prenatal development, histochemistry and innervations of porcine muscle. *Journal of Animal Science* **35**, 343-354.

Tabaries, S, J Lapointe, T Besch, M Carter, J Woollard, CK Tuggle and L Jeannotte (2005) Cdx Protein Interaction with Hoxa5 Regulatory Sequences Contributes to Hoxa5 Regional Expression along the Axial Skeleton. *Molecular and Cellular Biology* **25**: 1389–1401.

Tang, Z, Y Li, P Wan, X Li, S Zhao, B Liu, B Fan, M Zhu, M Yu and K Li (2007) LongSAGE analysis of skeletal muscle at three prenatal stages in Tongcheng and Landrace pigs. *Genome Biology* **8**, R115.18.

Te Pas, MFW, AAW de Wit, J Priem, M Cagnazzo, R Davioli, V Russo and MH Pool (2005) Transcriptome expression profiles in prenatal pigs in relation to myogenesis. *Journal of Muscle Research and Cell Motility* **26**, 157-165.

Te Pas, MFW, I Hulsegge, A Coster, MH Pool, HH Heuven and LLG Janss (2007) Biochemical pathways analysis of microarray results: regulation of myogenesis in pigs. *BMC Developmental Biology* **7**, 1-15.

Vodovar, N, F Desnoyers and AC Francois (1971) Origine et evolution des adipocytes mesenteriques du porcelet avant la naissance. Aspect ultrastructural. *J Micros* **11**: 265-284.

White, JD, C Rachel, R Vermeulen, M Davies and MD Grounds (2002) The role of p53 in vivo during skeletal muscle post-natal development and regeneration: studies in p53 knockout mice. *International Journal of Developmental Biology* **46**: 577-582.

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

Yang, W, H-X Yan, L Chen, Q Liu, Y-Q He, L-X Yu,S-H Zhang, D-D Huang, L Tang, X-N Kong, C Chen, S-Q Liu, M-C Wu and H-Y Wang (2008) Wnt/B-Catenin Signaling Contributes to Activation of Normal and Tumorigenic Liver Progenitor Cells. *Cancer Research* **68**: 4287-4295.

Yu, YH, BH Liu, HJ Mersmann and ST Ding (2006) Porcine peroxisome proliferator-activated receptor {gamma} induces transdifferentiation of myocytes into adipocytes. *Journal of Animal Science* **84**:2655-2665.

Zhang, DX, K Li, B Liu, ZM Zhu, XW Xu, SH Zhao, M Yerle and B Fan (2006) Chromosomal localization, spatio-temporal distribution and polymorphism of the porcine tripartite motif-containing 55 (TRIM55) gene. *Cytogeneics Genome Research* **114**:93B. doi: 10.1159/000091936.

Zhao, SH, D Nettleton, W Liu, C Fitzsimmons, CW Ernst, NE Raney and CK Tuggle (2003) Complementary DNA macroarray analyses of differential gene expression in porcine fetal and postnatal muscle. *Journal of Animal Science* **81**: 2179–2188.

CHAPTER III

A comparative analysis of microarray data set based on different protocols

Abstract

Applications of different methodological studies in microarray are subject to multiple sources of experimental variations. This chapter aimed to evaluate the application of two different protocols for data normalization and assessment of genes differentially expressed and compare with a previous microarray analysis (first chapter). The same data from the 13 slides were experimentally designed to proceed direct comparisons independently for each contrast and submitted to a background correction method called normexp and two within-array normalization methods were tested: Robust Spline and Loess. A trade-off among the results was observed when comparing the two current protocol analyses tested after statistical analyses based on the empirical Bayes method. The current protocols detected genes as DE only in two out of the four contrasts under evaluation (C40vsP40, P40vsP70, C70vsP70 and C40vsC70). The Robust Spline normalization was more robust to detect genes differentially expressed in the contrast between both genetic groups (Piau local breed and the crossbred YorkshirexLandrace) at 40d of gestation, while the contrast between both ages (40 and 70 days of gestation) in the Piau breed presented genes with significant FDR values more variable between both methods. The clustering analysis was able to represent genes DE with similar FDR values among the three protocols evaluated with functionalities designated to the proper contrasts in context (C40vsP40 or P40vsP70). The P40vsP70 contrast presented more number of sub clusters attempting to joint genes closely related in terms of FDR values, suggesting that age-specific genes are more sensitive to different normalization (and protocols) methods applications. The results strengthened the idea that the Robust Spline method was able to identify more genes differentially expressed (FDR ≤ 0.05) related with important GO terms involved with the muscle development process than the Loess method. The normexp background correction method is advisable, being clearly more accurate than the traditional subtraction method in the present study. The *q-value* FDR method is too flexible to detect genes differentially expressed in small microarray experiments, while the BH (Benjamini and Hockberg) FDR method can be too restrictive for that – although pointing out genes DE with higher fold change values and with similar functionalities related to the myogenesis process in context. The comparisons among protocol analyses for microarray data are necessary for cleaning and improving the quality of the measures of gene expression.

Keywords: normexp, robust spline, false discovery rate

Resumo

Aplicações de diferentes metodologias de análise de micorarranjos são sujeitas a múltiplos fatores de variações experimentais. Este capítulo avaliou aplicações de dois diferentes protocolos para normalização de dados e detecção de genes diferentemente expressos e comparou com uma prévia análise de microarranjos. Os mesmos dados gerados a partir de 13 lâminas de microarranjos foram experimentalmente organizados a fim de proceder comparações diretas independentes para cada contraste e foram submetidos a um método de correção de *background* chamado *normexp* e dois métodos de normalização dentro de lâminas foram testados: *Robust Spline* e *Loess*. Resultados conflitantes foram observados quando comparado os dois novos protocolos de análises testados baseando-se no método estatístico *empirical Bayes*. Os protocolos detectaram genes diferentemente expressos (DE) apenas em dois dos quatro contrastes sobre avaliação (C40vsP40, P40vsP70, C70vsP70 e C40vsC70). O método de normalização *Robust Spline* foi mais efetivo para detector genes DE no contraste entre ambos os grupos genéticos (raça local Piau e o cruzado YorkshirexLandrace) aos 40 dias de gestação, enquanto o contraste entre ambas as idades gestacionais (40 e 70 dias) dentro da raça Piau apresentou genes com valores sinificativos de FDR mais variáveis entre ambos os métodos de normalização. A análise de agrupamento foi capaz de representar genes DE com valores de FDR similares entre os três protocolos avaliados com funcionalidades apropriadamente designadas aos contrastes em questão (C40vsP40 or P40vsP70). O contraste P40vsP70 apresentou um maior número de subgrupos na tentativa de agrupar genes similares em termos de valores de FDR, sugerindo que genes idade gestacional–específicos são mais sensíveis a diferentes métodos de normalização (e protocolos). Os resultados enfatizam a idéia de que o método *Robust Spline* foi capaz de identificar um maior número de genes como DE (FDR $\leq$0.05) relacionados a importantes funções biológicas (*Gene Ontology*) envolvidas no processo de desenvolvimento muscular do que o método Loess. O método de correção de *background normexp* é recomendado, sendo nitidamente mais acurado do que o método tradicional de subtração no presente estudo. O método de FDR *q-value* é muito flexível para detector genes como DE em pequenos experimentos de microarranjos, enquanto o método *BH* pode ser muito restritivo – apesar de ser capaz de identificar genes DE com maiores valores de expressão relativa e funcionalmente relacionados ao processo de miogênese em contexto. Comparações entre protocolos de análises para dados de microarranjos são necessários para melhorar a qualidade e representar de forma mais clara as medidas de expressão gênica.

Palavras chave: *normexp, robust spline,* taxa de falsas descobertas

1. Introduction

The main issue in microarray experiment is data analysis and the consequent extraction of biological knowledge. Transcriptome analysis itself is complicated by multiple factors such as the limited number of possible experiment replications - which is always lower than the number of variables, i.e. genes under investigation (Draghici et al., 2001) - and the limited understanding of gene regulation and gene product function in different conditions or systems. Moreover, the large multiplicity problems in which thousands of hypotheses are tested simultaneously within one experiment (Dudoit et al., 2002) is another tricky characteristic in microarray analyses.

It is clear that any microarray data analysis can be summarized in four main steps, which can be completed using different computational tools and/or statistical methods: a) Quality control; b) Data pre-processing; c) Differential expression detection and d) Biological knowledge extraction. Identify any specific software or methodology that is globally accepted by the scientific community as the gold standard for microarray data analysis is being difficult but extensively discussed (Fujita et al., 2006; Workman et al., 2002) due to poor overall reliability in comparisons between different experiment designs and different platforms (Pham et al., 2006; Pedotti et al., 2008). According to Verducci et al. (2006), the lack of a robust and reliable data analysis platform represents the single most important limiting factor in microarray analyses.

Different methodological studies in microarray drastically affect the results and consequentially the interpretations about transcriptional profiles. Moreover, any microarray experiment procedure and analysis is subjected to multiple sources of experimental variation itself. Some of these variations are considered systematic and may be explicitly corrected taking account background corrections methods and through data normalizations with the objective of cleaning and improving the quality of the measures of gene expression (Soler et al., 2004).

Thus, this chapter aims to evaluate the application of different protocols for data normalization and assessment of genes differentially expressed and compare with a previous one (chapter 1). The same data from the 13 slides, now experimentally designed to account for direct comparisons independently

within each contrast, were submitted to a background correction method called normexp (Ritchie et al., 2007) and two distinct within-array normalization methods (Robust Spline and Loess).

2. Review

In order to accurately and precisely measure gene expression changes, it is important to take into account the random (experimental) and systematic variations that occur in every microarray experiment. Even though such systematic biases can be comparatively small, they may be confounding when searching for subtle biological differences. In that context, considering two-channel microarrays, the term normalization refers to the process of removing such artefactual bias (Yang et al., 2002), which can be intensities of both dyes (Cy3 and Cy5, in two-channel microarray) between samples on the same slide and also between slides inherent to the experiment (Yang and Thorne, 2003).

2.1. Background correction

Within the data pre-processing step, the pixel intensities from the Cy3 (green, G) and Cy5 (red, R) images are measured and the image analysis software returns foreground and background intensities for each spot. Sources of variation such as non-specific binding of labeled sample to the array surface, processing effects such as deposits left after the wash stage, or optical noise from the scanner can be designated as background (Bengtsson & Bengtsson, 2006). While the background always co-exists with the overall measure of the intensity of the spot (foreground), these systematic sources should be removed, once systematic effects resulting from biological process under study are the real interest.

Remove non-specific signal from the total intensity in principal is an unbiased estimator of the true signal, but proceed or not a 'background correction' is still an arguably decision. A simple background subtraction such that $R = Rf - Rb$ and $G = Gf - Gb$, can produces negative intensities whenever the background intensity is larger than the foreground intensity, leading to missing log-ratios, sometimes for a substantial proportion of probes on an array (Beißbarth et al., 2000; Kooperberg et al., 2002; Bilban et al., 2002). However, the normexp method (Ritchie et al., 2007), a model-based

adjustment provided by the limma software (Smyth, 2005), has being successfully applied in different studies (Weniger et al., 2007; Peart et al., 2005; Gilad et al., 2006; Oshlack et al., 2007). This background method proposed, observes the pixel intensities as the sum of two random variables, one normally distributed and the other exponentially distributed, representing background noise and signal, respectively (Silver et al., 2008).

2.2. Loess normalization method

The microarray data can be represented by an M versus A plot, or MA plot (Dudoit et al., 2002), where the log-ratios are given by M = log2(R/G) and the average log-intensity by A = log2 √RG. Imbalance in the red (R) and green (G) intensities is usually not constant across the spots within and between arrays, and can vary according to overall spot intensity A, location on the array, plate origin, and possibly other variables (Hartemink et al., 2001).

Yang et al., (2002) proposed an upgrade of the Global normalization method which assumes that the red and green intensities are related by a constant factor, i.e. R=kG, and the center of the distribution of log ratios is shifted to zero. Aiming to consider intensity-dependent bias frequently detected in microarray experiments, the method named Local-Weighted Regression and Smoothing Scatterplot – lowess (or Loess, according to Cleveland & Devlin, 1988), was implemented in the statistical software package R to perform a local A-dependent normalization (Equation 1). Actually, the Loess method is derived from the S statistical function Lowess (Cleveland, 1979), which uses a locally weighted least squares estimate of a regression fit. In simple terms, we take continuous sections of the data and fit a linear (or quadratic) regression line to that area of data.

$$log_2 \frac{R}{G} \rightarrow log_2 \frac{R}{G} - c(A) = log_2 \frac{R}{[k(A)G]} \qquad \text{Equation 1}$$

Where c(A) is the Loess fit to the MA-plot and k in a constant factor relating the red (R) and green (G) intensities.

The Loess normalization method is a non-parametric method to the regression of the M on A values which can also fits different intensities- dependent curves to different regions of the array (Print-tip groups) (Equation 2).

$$log_2 \frac{R}{G} \rightarrow log_2 \frac{R}{G} - c_i(A) = log_2 \frac{R}{[k(A)G]} \qquad \text{Equation 2}$$

Where c_i is the Loess fit to the MA-plot for the *i*th grid only (each print-tip group).

The Loess method assumes that in each intensity interval either the majority of genes are non-differentially expressed or the numbers of "up"- and "down"-regulated genes are equal (Zhao et al., 2005).

2.3. Robust Spline normalization method

It is believed now a day that others adaptive non- linear methods more robust can generate a better normalization within-array analyses (Soler et al., 2004).

According to Smyth (2004) and stated by others (Rainer et al., 2006), the Robust Spline Normalization (RSN) is a function which implements regression splines in place of the Loess curves and uses empirical Bayes (eBayes) ideas to shrink the individual print-tip curves towards a common value to normalize the M-values for each microarray. This feature allows the technique to introduce less noise into good quality arrays with little spatial variation while still giving good results on arrays with strong spatial variation.

In a real sense, splines are an evolution of classical parametric inference and bridge the gap between parametric and nonparametric methods (Wegman & Wright, 1983). Splines methods aim to fit a smooth curve between specified points named knots.

More to the point, it is desirable in a statistical framework to create a type of smoothing spline that could pass near, in some sense, to the data but not be constrained to interpolate exactly. The regression spline method satisfies more than the continuity conditions similar to other simplest splines methods, it does not necessarily minimize the curvature norm. This special feature is interesting when dealing with noisy data, like normally seen in microarray experiments.

Basically, a basis function can be assumed to fit functions (based on a settle time *t*) that are non-linear with respect to the original input set:

$$\tilde{y}_t = \beta_0 + x_{1t}\beta_1 + x_{2t}\beta_2 + \varepsilon_t$$

Where,

$x_{1t} = \log(t)$ and $x_{2t} = t$.

To approximate non-linear functions using local linear functions we can use for example two interior knots (τ_1, τ_2) to generate three windows on *t* , e.g.:

$$\{0 < t \leq \tau_1;\ \tau_1 < t \leq \tau_2; t > \tau_2\}$$

In this way, the conditional expectation function will be:

f_t=

$$\begin{cases} \beta_1 + \beta_2 t \ \ if \ \ t \leq \tau_1 \\ \beta_3 + \beta_4 t \ \ if \ \ \tau_2 < t \leq \tau_1 \\ \beta_5 + \beta_6 t \ \ otherwise \end{cases}$$

Two restrictions to perform separate linear regressions within each interval, will assure that the locally fitted linear functions will match at the knots (i.e., the conditional expectation will be continuous):

$$\beta_1 + \beta_2 \tau_1 = \beta_3 + \beta_4 \tau_1 \quad \text{and} \quad \beta_3 + \beta_4 \tau_2 = \beta_5 + \beta_6 \tau_2$$

So that, the fitted function is continuous at the two interior knots that define the three windows to fit a linear approximation- the conditional expectation function. If we do so, the effective number of parameters in the model will be 4, instead of 6:

$$x'_i\beta = \beta_1 + \beta_2 t + \beta_3(t - \tau_1) + \beta_4(t - \tau_2)$$

According to Smith and Kohn (1996) the Bayesian approach robustify linear regression by modeling the errors as a mixture of normal with greater local adaptability to the shape of the function.

2.4. Quantile normalization method

The effect of location normalization is to center log-ratios around zero by accounting for intensity- and spatially-dependent bias. In addition, it is important to consider scale normalization, since large scale differences between multiple slides can lead some slides giving undue weight to an average of log-ratios across slides.

Quantile normalization is to ensure that the intensities have the same empirical distribution across arrays and across channels. This method was proposed by Bolstad et al., (2003) in single channel arrays and by Yang and Thorne (2003) for two-color cDNA arrays.

This method extends the idea of normalizing for equivalent medians or quartiles of the single-channels by requiring every quantile across channels are equivalent (Bolstad et al., 2003) and thus forcing each channel to share a common distribution. This implies that we can give each array the same distribution by taking the mean quantile and substituting it as the value of the data item in the original dataset.

2.5. False Discovery Rate

Because microarray experiments generate large multiplicity problems in which thousands of hypotheses are tested simultaneously within one experiment, an adjustment of the calculated P-values should be performed in statistical analysis of each gene to discover the truly differentially expressed ones (Benjamini et al., 2005).

Considering the rate of judge each gene *i (i=1 and 2)* as differentially expressed (reject the null hypothesis) when there is truly no significant difference, as α=0.05. Otherwise, admitting the individual possibility to each hypothesis test to be correct, when detecting both genes as differently expressed equal to $(1-\alpha)^2$, imagine how much would be the probability to commit the mistake? Almost twice as much:

$1-(1-\alpha)^2 = 0.0925$.

The FDR (False Discovery Rate) procedure came abroad when researchers started overlooking various kind of multiplicity, because reporting's tended to exaggerate treatment differences when single-comparisons were done between multiple treatments, proposing thus, to account all the erroneous rejection and not only the question whether any error was made. Benjamini and Hockberg (1995) proposed a linear set-up procedure commonly named BH for controlling the expected proportion of falsely rejected hypotheses – the most traditional FDR. An unguarded use of single-inference when pursuing multiple inferences procedures results in a greatly increased false positive rate.

Consider testing $H_1, H_2, \ldots, H_m$ based on the corresponding p-values $P_1, P_2, \ldots, P_m$. Let $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(m)}$ be the ordered p-values, and denote by $H_{(i)}$ the null hypothesis corresponding to $P_{(i)}$. Define the following Bonferroni-type multiple testing procedure:

Let *k* be the largest *i* for which $P_{(i)} \leq \frac{i}{m} q^*$

Then, reject all $H_{(i)}$, $i = 1, 2, \ldots, k$. (1)

Where m_o = total number of true null hypotheses

$M_1 = m - m_o$= total of non-true null hypotheses or number of truly alternative features.

For independent test statistics and for any configuration of false null hypothesis, this is the procedure to control the FDR at q^*. This theory follows the lemma:

For any $0 \leq m_o \leq m$ independent p-values corresponding to true null hypothesis and for any value that the $m_1=m-m_o$ p-values corresponding to the false null hypothesis can take, the multiple testing procedure defined in (1) satisfies the inequality:

$$E\left(Q \middle| P_{m_o+1} = p_1, \ldots, p_m = p_{m_1}\right) \leq \frac{m_o}{m} q^* \qquad (2)$$

Now, suppose that $m_1=m-m_o$ of the hypothesis is false. Whatever the joint distribution of p_1", …, p_m1" which corresponds to the false hypothesis is integrating inequality (2) above, we obtain:

$$FDR = E(Q) \leq \frac{m_o}{m} q^* \leq q^*;$$

and the FDR is controlled.

According to Reiner et al. (2003) substantial increase in power is already gained when the p-values are estimated by resampling, and then used in the BH procedure (Benjamini and Hochberg, 1995). However, a lot of adaptative procedures have been proposed to deal with for positively dependent test statistics as well (Benjamini and Yekutieli, 2001; Benjamini et al., 2001; Yekutieli and Benjamini, 1999).

The R package provides suitable methods to adjust P-values regarding this multiple hypothesis testing problem.

The practical goal of genomewide studies is to estimate as many significant feature (e.g. genes) in the genome as possible - simultaneously testing thousands of features - while incurring a relatively low proportion of false positives (Storey and Tibshirani, 2003). Aiming to provide an evolution of the commonly used FDR in a genomic statistics context, these authors discussed the basic differences between concepts of false positive rate (rate that truly null features are called significant) and false discovery rate (rate that significant features are truly null) and propose that the *q-values* (proposed FDR)

directly provide a meaningful measure among the features called significant because measure the proportion of significant features that turn out to be false leads.

In resume, the *q value* of a particular feature (gene) can be described as the expected proportion of false positives among all features as or more extreme than the observed one. The *q value* has a special probabilistic relationship to the *p value* (yielding the origin of its name).

Let P_i be the p value of the ordered gene i. R_i is the total number of rejected genes whose p values are less than the threshold t = P_i and $\hat{m}_o$ is an estimate of the total number of non differentially expressed genes, m_o.

$$FDR = q\ value = \frac{\hat{m}_o\ P_i}{R_i}$$

Basically, Benjamini & Hochberg (1995) propose the false discovery rate concept and provide the first step-wise *p value* method to control it. Storey & Tibshirani (2003) generally defines the *q value* and initially studies its properties in a Bayesian context.

It is known that the methods to control the FDR in microarray experiments cited above are widely used and are a less stringent condition than the family wise error rate (FWER) – wich control single comparisons problems. However, when each of them is more powerful and properly in different experiment conditions, is still a questionable issue.

3. Material and Methods

3.1. Experiment design

Based on the same Swine Protein-Annotated Oligonucleotide Microarray (Pigoligoarray; www.pigoligoarray.org) - which includes 20,400 70-mer oligonucleotides - used to proceed the previous trascriptome profiling evaluation of the *Longissimus dorsi* muscle of two genetic groups of pigs at two

prenatal ages - the row data (.gpr files) from the 13 slides analyzed on the Axon 118 GenePix® 4000B scanner (Molecular Devices), were used in this study.

At this time, the 13 slides were separated in four independent experiments in which: 3 slides represented the breed comparison at 70d of gestation, 3 slides represented the breed comparison at 40d of gestation, 3 slides represented the age comparison for the YL crossbred and 4 slides represented the age comparison for the Piau breed (Figure 1).

Each slides set organized in a text file (.txt) named *target* was uploaded in to R package (Gentleman et al., 2004) limma (R 2.3.1, Smyth, 2005), followed by reading images step and recognition of the *gal file*.

3.2. Data processing

The normexp background correction method and the subtract background methods were tested in to the slides set and represented in box-plots.

After background correction, the measurements from the Cy5 channel (Rp) and the measurements from the Cy3 channel (Gp) for each probe *p* were converted to M and A values for normalization. The entire parameters become additive, once on the log scale.

At this time the two normalization whiting-array methods (Robust Spline and Loess) were separately tested in the current data set and the quantile method was applied to proceed the normalization between-array.

For the comparisons of interest, moderated t-statistics that uses an empirical Bayes method were calculated (Smyth, 2004). Differentially expressed genes were identified using the FDR corrected p-value ≤ 0.05 (Benjamini and Hochberg, 1995).

The contrasts (*C*) were calculated based on normalized logarithmic expression values (E) as $C = (E(BA_r) - E(BA_t))$; where *r* indicates the reference treatment, *t* indicates the tested treatment for genetic groups (B) and prenatal ages (A).

Four separated lists of genes observed as differentially expressed between the four contrasts for each normalization methods applied were generated and exported to an excel file to perform a comparative analysis. The *macro* DigDB (http://www.digdb.com) recognized in *excel* was used to merge all the excel files to the "array annotation" file. This file contains identification information (ID), HUGO Gene Nomenclature Committee (HGNC) annotation and GO annotation of the 18,524 targets represented in the slide set analyzed (Steibel et al., 2009).

3.3. Clustering Analysis and Gene Ontology Analysis

In order to explore the comparisons between protocol analyses a clustering analysis was proposed to identify genes that independently of the protocol applied, presented significant or similar fold change values.

The clustering analysis based on the centroid method was applied in SAS (9.2) to separate groups of genes based on similar FDR values detected above each normalization procedure. This method assigns each item to the cluster having nearest centroid (means), based on the Euclidian distance (Johnson & Wichern, 1998). The RMSSTD (Root Mean Square Standard Deviation) parameter measures the homogeneity of the clusters suggested. Plotting this values against the number of clusters generated permits the identification of the maximum curvature, which is the best number of clusters to be considered (Sharma, 1996).

For classification of microarray data between developmental ages or between genetic groups, the DAVID software (Database for Annotation, Visualization and Integrated Discovery; http://david.abcc.ncifcrf.gov/; Dennis et al. 2003; Huang et al. 2009) was used in order to determine functional classifications based on gene ontology categories.

4. Results

The box-plots presented in the figure 1 (Appendix 3) revealed that the normexp method resulted in more striking resemblance between arrays than the subtract background method did.

Thus, the MA-plots comparing both normalization procedures applied (Robust Spline and Loess), after normexp background correction method, were presented in the figure 2 (Appendix 3). Each slide was represented by three scatter-plots showing the respective dispersion of the log-ratios against the log-intensity of each oligo before the normalization procedure, after normalization according to the Robust Spline method and after normalization according to the Loess method, respectively. According to the shape of the colorful curves, representing the trend of expression of each print-tip group of the slide, the Robust Spline normalization was able to centered them more closely to M=0.

4.1. Identifying genes differentially expressed

The design proposed for the 13 slides arranged in four independent experiments is shown in the figure 1. The arrows identified the slides, and the respective group comparison labeled (Cy3 or Cy5) is in accordance to the sense of the arrows. After proceding the statistical analyses in each of the four experiments separately normalized with Robust Spline and Loess methods, the results were compared.

Interestingly, the first experiment which compared both genetic groups at 70 d of gestation (C70 vs P70) did not show any differentially expressed (DE) gene (FDR $\leq$0.05), independently of the normalization method applied. Slight difference in the fold change and FDR values comparing the Robust Spline and the Loess methods was detected and the smallest FDR value found in the "Loess list" was 0.098135, against 0.081702 in the "Robust Spline list".

The second experiment, which compared both genetic groups at 40 d of gestation, presented a first sub list with 221 oligos with HGNC information differentially expressed in both normalization methods applied. A second sub list was formed by 110 exclusively differentially expressed when applying the Robust Spline method and a third sub list represented for only 13 oligos exclusively differentially expressed when applying Loess normalization (data not shown).

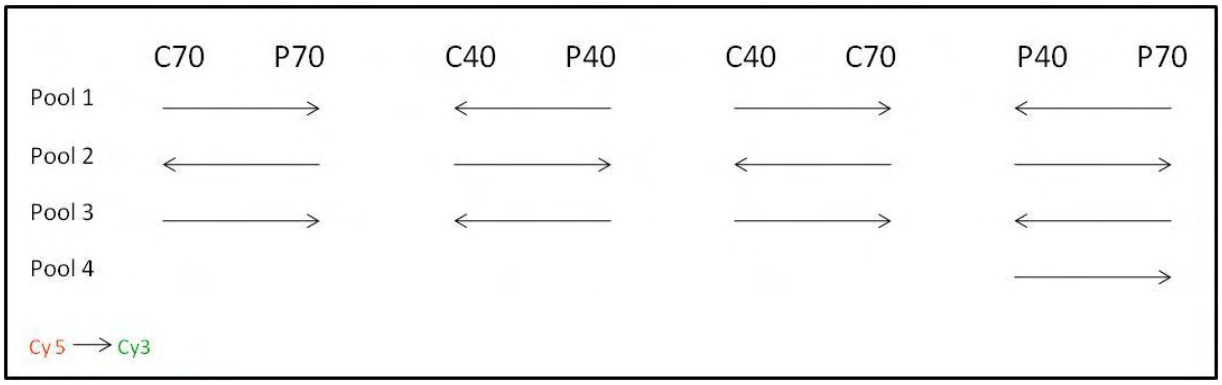


Figure1. Experiment design. 13 microarray slides separated in four independent experiments, respectively: Experiment 1 (C70 vs P70) Breed comparison at 70 d of gestation; Experiment 2 (C40 vs P40) Breed comparisons at 40 d of gestation; Experiment 3 (C40 vs C70) Age Comparison in the YL crossbred; Experiment 4 (P40 vs P70) Age comparison in the Piau breed. The pool of samples used in each slide and the arrow sense identification between dyes are represented.

The third experiment involving age comparisons in the crossbred Yorkshire-Landrace (C40 vs C70), presented only five genes DE (FDR≤0.05) (MYL7, *Myosin regulatory light chain 7;* TPM1, *Tropomyosin 1 alpha chain*; FABP4, *Fatty acid-binding protein, adipocyte* and ART3, *Ecto-ADP-ribosyltransferase 3 precursor*) when proceeding the Robust Spline normalization, while only one gene was DE (FDR≤0.05) in the Loess list; the MYL7 gene (*Myosin regulatory light chain 7*).

The last experiment representing the comparisons between both prenatal ages in the Piau breed (P40 vs P70), revealed 148 genes DE in both normalization methods applied. Another 34 genes were exclusively DE when applied the Robust Spline method, and others 47 exclusively DE when applied the Loess method (data not shown).

Interestingly, 34 oligos with HGNC information not detected as DE in the previous analysis (first chapter) in the contrast between both breeds at 40 d of gestation, were now detected as DE when one or the other normalization method were applied after normexp background correction (Table 1). In the same way, another 37 genes were now detected as DE in the comparisons between both prenatal ages in the Piau breed (Table 2). The Robust Spline method detected more genes DE (FDR≤0.05) than the Loess

method in both contrasts, even though the contrast between ages in the Piau breed presented more variable FDR values between both methods.

Table 1. New genes detected as differentially expressed in the contrast between both genetic groups at 40 d of gestation (Piau breed is the reference) in each type of normalization method applied after background correction (normexp). Robust Spline (RS) or Loess (L). Name of the genes (HGNC) and respective fold change value, p-value and FDR value detected for each normalization method.

NAME	FC RS	PValueRS	FDR RS	FC LB	PValueLB	FDR LB
APCDD1	-3.395891342	0.000651	0.044315367	-3.155759	1.41E-03	0.059687762
COX11	2.817204145	7.49E-05	0.037601612	2.477338	8.99E-04	0.051640399
CPM	-1.688417586	0.000577	0.042751163	-1.749364	3.77E-04	0.043403998
CRCP	1.801259766	0.000266	0.037601612	1.921894	2.42E-04	0.040428729
CST3	-1.6559363	0.000629	0.044236537	-1.704469	1.31E-03	0.057663275
DNER	-2.279306558	0.000961	0.048217568	-2.283327	7.61E-04	0.04902451
EEF2	2.336231562	0.000143	0.037601612	2.582488	8.73E-05	0.040428729
EFEMP1	-1.955676543	0.000832	0.047008111	-1.854927	8.51E-04	0.050958682
ESPL1	-1.971762263	0.000662	0.044315367	-1.949502	1.03E-03	0.052780756
FABP4	-2.029848604	0.000744	0.04579017	-2.095723	3.79E-04	0.043403998
HIF3A	-2.632851162	0.000107	0.037601612	-2.639353	1.16E-04	0.040428729
HINFP	-2.122434211	0.000314	0.037601612	-2.201599	5.93E-04	0.047606612
HINFP	-1.770177212	0.000539	0.04233298	-1.764359	7.38E-04	0.048726859
HINFP	-2.266282265	0.000814	0.046604655	-2.29808	1.05E-03	0.053581952
HLA-G	-1.708899058	0.000377	0.039103208	-1.741566	3.92E-04	0.043403998
HSPA1B	-7.500024516	2.51E-05	0.037601612	-7.96001	3.60E-05	0.040428729
KCNK18	-2.228589555	1.67E-03	0.059637272	-2.324003	7.42E-04	0.048726859
KIF13B	-1.674522596	0.000654	0.044315367	-1.640962	9.00E-04	0.051640399
KIF5A	-2.500837664	1.22E-03	0.052902171	-2.600742	4.62E-04	0.044719466
LYCAT	-1.592686247	0.00082	0.046668067	-1.602851	9.95E-04	0.052365952
MED14	-1.613252774	0.000794	0.046604655	-1.51842	2.22E-03	0.069380082
mt-Nd1	1.792786728	0.000365	0.039103208	1.693909	9.95E-04	0.052365952
mt-Nd5	2.877786063	5.59E-05	0.037601612	2.837874	2.18E-04	0.040428729
NPDC1	-2.044799204	0.000931	0.048004064	-2.068258	1.01E-03	0.052735129
PCSK1N	-1.982291662	0.000124	0.037601612	-2.043952	1.11E-04	0.040428729
PPP1R14A	-1.722648174	0.000896	0.047182986	-1.773896	1.60E-03	0.062062549
PTPN18	-1.60201353	0.000805	0.046604655	-1.606299	1.27E-03	0.057663275
S100A6	-1.770235587	0.000373	0.039103208	-1.696544	6.10E-04	0.047606612
SEPP1	1.641660661	1.23E-03	0.052902171	1.718125	6.58E-04	0.048588946
SH3BP5	-1.576551231	0.001042	0.049367843	-1.597217	9.56E-04	0.052132706
SRrp40	1.730981465	0.000875	0.047101134	1.627404	1.45E-03	0.060264005
TUBB2A	1.571686879	0.001045	0.049367843	1.611722	8.78E-04	0.051152883
UROD	1.628649102	0.000763	0.046308975	1.61551	9.21E-04	0.051860275
ZFP36L2	-2.016729352	0.000333	0.037601612	-2.004839	2.94E-04	0.041850744

FDR ≤0.05: Genes differentially expressed. In red, genes with FDR > 0.05. Same HGNC's means different oligos ID's.

Table 2. New genes detected as differentially expressed in the contrast between two prenatal ages (40 and 70 d) in the Piau breed (70 is the reference) in each type of normalization method applied after background correction (normexp). Robust Spline (RS) or Loess (L). Name of the genes (HGNC) and respective fold change value, p-value and FDR value detected for each normalization method.

HGNC	FC RS	PValueRS	FDR RS	FC LB	PValueLB	FDR LB
ACTC1	-1.839859458	8.93E-04	0.062341404	-1.812741936	3.08E-05	0.012269708
AK1	-1.605004266	2.11E-03	0.089217808	-1.686268867	2.96E-04	0.036969013
C14orf122	-1.578639191	2.86E-04	0.037524657	-1.608963325	7.59E-04	0.057221413
C22orf16	-2.168459115	6.43E-05	0.01825973	-2.006468408	5.02E-04	0.047952981
C4orf8	-1.455117175	3.02E-04	0.038537775	-1.472129036	9.00E-04	0.062017529
C6orf173	1.405320029	4.37E-04	0.045314828	1.407649628	1.15E-03	0.070718374
CCDC21	-1.592203288	9.76E-05	0.022548807	-1.578705449	3.33E-04	0.038535372
CCNB1	1.516898938	5.78E-04	0.050467924	1.552145119	1.70E-04	0.027472841
CDH11	2.240368478	1.84E-04	0.030073965	2.27797521	1.17E-04	0.023357703
CDH15	-1.368034431	1.32E-03	0.071503785	-1.426600914	5.57E-04	0.049468133
CDKN2AIPNL	1.443360583	2.93E-04	0.038212594	1.411599682	7.92E-04	0.058022619
COL8A1	1.410191526	3.49E-04	0.040597911	1.427094403	7.25E-04	0.055488316
COX6A2	-1.64022755	5.40E-04	0.048813451	-1.697527465	4.61E-04	0.046417232
DNASE1L1	-1.579729691	2.77E-04	0.037080088	-1.596106705	2.84E-04	0.035958335
EHBP1L1	-1.564264688	1.58E-04	0.029130751	-1.576161822	2.87E-04	0.036097455
FHL1	-2.160694067	1.39E-05	0.008760187	-2.178756872	1.40E-04	0.025546791
FST	-1.539921123	2.72E-04	0.037067389	-1.49501525	1.74E-03	0.086458098
GOLGA4	-1.590188735	2.33E-03	0.094345821	-1.620851696	5.10E-04	0.04850572
HERC2	-1.480493715	1.70E-04	0.029576609	-1.489159726	8.37E-04	0.059927015
HSPB1	-1.603044922	1.56E-04	0.029130751	-1.524216354	8.80E-04	0.061248152
LMO4	1.407175961	4.51E-04	0.046083328	1.393516509	1.22E-03	0.073737159
MACF1	1.35920327	4.91E-04	0.047529603	1.515028712	3.39E-04	0.038535372
MMP23A	-1.529304151	3.67E-04	0.041361632	-1.576911973	1.57E-04	0.027263929
MPPED2	2.337432888	4.39E-04	0.045314828	2.439759164	1.00E-04	0.022148872
MYL2	-2.514746887	3.25E-05	0.012286959	-3.927814863	4.06E-06	0.007888568
OSTN	2.901564517	5.25E-04	0.048344534	2.855742424	6.90E-04	0.053994356
PLS3	1.493809505	3.50E-04	0.040597911	1.564965658	7.55E-04	0.057108104
PNRC1	-1.885393732	8.49E-05	0.021078054	-1.962761911	5.67E-05	0.016317058
PRRX1	1.667734894	7.90E-04	0.057709146	1.735029593	5.81E-04	0.049810888
RBM24	-1.47067528	4.44E-03	0.13269584	-1.501064568	5.31E-04	0.049135759
S100A13	-1.349622286	5.09E-04	0.047777655	-1.343009527	1.10E-03	0.069229011
ST3GAL4	-1.451117375	1.04E-03	0.066213361	-1.429213091	5.33E-04	0.049135759
SYPL1	-1.457551484	1.59E-04	0.029130751	-1.485594438	4.97E-04	0.047946968
TM6SF1	-1.940623763	1.63E-04	0.029576609	-1.777309103	3.93E-04	0.041606999
TNNT2	-1.491143505	3.19E-04	0.039218736	-1.652267987	3.40E-04	0.038535372
TSPAN6	1.474747606	1.52E-04	0.029130751	1.447262512	1.74E-03	0.086458098
UBL4A	-1.420073177	5.43E-04	0.048813451	-1.385178816	2.05E-03	0.093358942

FDR ≤0.05: Genes differentially expressed. In red, genes with FDR > 0.05. Same HGNC's means different oligos ID's.

4.2. Clustering analysis

Two out of the four contrasts, analyzed by the two different protocols analyses, presented relevant amount of genes differentially expressed. In order to compare these two contrasts with the results obtained by the previous mixed model analysis, a clustering analysis was applied to help further speculations. Once the FDR represents the adjusted value which designate a gene as differentially expressed or not, among others submitted to the same protocol analysis, the clustering procedure generated a "main list" of genes suggested to be more similar according to their FDR values (criteria) among the three protocol analysis proposed for each contrast (C40vsP40 and P40vsP70).

For the contrast comparing both genetic groups at 40 d of gestation (C40vsP40), the analysis suggested a separation of the whole data in 500 different clusters. This conclusion was taken based on a graph, according to the maximum curvature of the proposed clusters distributed in terms of numbers of genes in each (not shown). Among these 500 clusters, the first one proposed presented the lowest average of FDR values in each of the three protocol analysis (previous Loess normalization method with no background correction, current Robust Spline and Loess normalization methods background corrected): 0.01202, 0.050006 and 0.055627, respectively. The number of oligos with HGNC information included in this cluster was 536. When proceeding another clustering analysis within this list of 536 genes, 50 sub clusters were suggested, but four of them (1st, 2nd 3rd and 24th), presented the lowest FDR averages among the three analysis (Table 3 and 4).

Table 3. Four sub clusters more informative within the 50 suggested for the contrast between both genetic groups at 40 d of gestation in terms of FDR values obtained after analysis. Number of the cluster, number of genes in the cluster (FREQ) and FDR average of the cluster respectively in the previous analysis (L) and current analysis background corrected before using the Robust Spline (RS) and Loess (LB) as normalization methods.

CLUSTER	_FREQ_	FRD L	FDR RS	FDR LB
1	94	0.00303	0.044278	0.050015
2	10	0.042764	0.039483	0.042746
3	128	0.004222	0.038379	0.041882
24	5	0.028547	0.037602	0.042539
	237			

Table 4. Name of the gene, respective FDR values and number of the sub cluster (C) more informative within the 50 suggested for the contrast between both genetic groups at 40 d of gestation in terms of FDR values obtained after statistical analysis.

NAME	FRD L	FDR RS	FDR LB	C	_NAME_	FRD L	FDR RS	FDR LB	C
ABHD14A	0.007713	0.04241	0.047607	1	CILP2	0.0437	0.042755	0.040429	2
ABI3BP	0.041835	0.037602	0.040429	2	CITED2	0.003082	0.037602	0.044291	3
ACMSD	0.003822	0.037602	0.040429	3	CNIH	0.001535	0.037602	0.047607	1
ADAM9	0.004383	0.039949	0.047607	1	COL1A1	0.002618	0.040504	0.047429	1
ALMS1	0.001572	0.048531	0.051031	1	COL27A1	0.006501	0.037602	0.040429	3
ANGPTL2	0.012738	0.039103	0.045032	3	COL4A1	0.047355	0.039103	0.042641	2
ANKMY2	0.003107	0.046309	0.05003	1	COL5A2	0.029393	0.037602	0.045032	24
ANKRD12	0.002406	0.037602	0.043404	3	COMMD10	0.00189	0.037602	0.040429	3
ANKRD37	0.003987	0.037602	0.040429	3	COX7C	0.003979	0.047099	0.047064	1
ANP32E	0.0011	0.044315	0.049335	1	CSDE1	0.00825	0.037602	0.040429	3
AP3M1	0.001817	0.048525	0.05186	1	CST6	0.030514	0.037602	0.040429	24
ARHGEF3	0.0011	0.047524	0.047429	1	CTSF	0.007215	0.037602	0.040429	3
ARID1A	0.001121	0.048842	0.052598	1	CUEDC1	0.005152	0.039103	0.040429	3
ARID1B	0.000787	0.037602	0.040429	3	CWF19L2	0.003987	0.042333	0.048589	1
ARL8B	0.001435	0.041847	0.048589	1	DAAM1	0.000971	0.037602	0.040429	3
ASNS	0.005152	0.04241	0.040429	3	DAB2	0.004061	0.039124	0.040429	3
ASPN	0.008626	0.037602	0.047607	3	DBC1	0.006982	0.037602	0.040445	3
ATIC	0.0011	0.047099	0.047607	1	DDIT4	0.013659	0.037602	0.045013	3
ATP5B	0.009248	0.037602	0.047678	3	DDX52	0.002681	0.042555	0.047607	1
ATP9A	0.00036	0.047524	0.053888	1	DKK2	0.04716	0.037602	0.040429	2
BAI2	0.002012	0.047063	0.051153	1	DNAJA1	0.001211	0.037602	0.040429	3
BAZ2A	0.000508	0.037602	0.040429	3	DNM1L	0.00118	0.040042	0.044291	3
BCHE	0.006203	0.040449	0.04539	3	DUSP1	0.03382	0.037602	0.040429	24
BCHE	0.0011	0.037602	0.042641	3	EGFLAM	0.003432	0.045537	0.056448	1
BLCAP	0.003964	0.044596	0.053888	1	EIF2S1	0.000169	0.037602	0.040429	3
BMPR2	0.005858	0.039103	0.044238	3	EPN2	0.003027	0.040504	0.044148	3
BNIP3	0.001926	0.037602	0.040429	3	FABP5	0.007173	0.037602	0.040429	3
BPTF	0.000971	0.037602	0.040429	3	FAM110B	0.0011	0.039124	0.047607	1
BTG2	0.004164	0.044596	0.047607	1	FAM114A1	0.009302	0.037602	0.051838	1
C10orf10	0.002475	0.037602	0.040429	3	FAM115A	0.011915	0.044237	0.044291	3
C10orf11	0.009295	0.037602	0.040429	3	FAM80B	0.004781	0.043762	0.050312	1
C10orf6	0.007233	0.041847	0.0522	1	FANCI	0.000787	0.048306	0.047607	1
C14orf14	0.001776	0.046605	0.045032	1	FAP	0.001244	0.037602	0.040429	3
C1QTNF4	0.011629	0.037602	0.040429	3	FLII	0.001435	0.037602	0.040429	3
C22orf28	0.00089	0.039556	0.047607	1	FLRT2	0.002467	0.046605	0.042641	1
CABYR	0.001067	0.037602	0.046744	1	FNDC3B	0.022726	0.037602	0.043404	24
CALU	0.000603	0.037602	0.043404	3	FOS	0.012222	0.040504	0.043404	3
CAP2	0.007948	0.037602	0.040429	3	FSTL1	0.002468	0.037602	0.042641	3
CAPZA1	0.0011	0.037602	0.040429	3	FYTTD1	0.001435	0.045537	0.049335	1
CAT	0.004014	0.037602	0.040429	3	GADD45B	0.003556	0.037602	0.040429	3
CCNG1	0.004524	0.037602	0.040429	3	GADD45B	0.007042	0.037602	0.040429	3
CCT5	0.006501	0.037602	0.040429	3	GALNT1	0.00698	0.045537	0.053888	1
CD164	0.001649	0.037602	0.040429	3	GALNT7	0.00209	0.044315	0.048589	1
CDO1	0.005577	0.037602	0.045641	3	GLG1	0.010334	0.037602	0.040429	3
CHD6	0.00951	0.037602	0.043404	3	GYG1	0.012832	0.041361	0.048589	3

FDR ≤0.05: Genes differentially expressed.

Table 4 cont.

NAME	FRD L	FDR RS	FDR LB	C	_NAME_	FRD L	FDR RS	FDR LB	C
H19	0.011692	0.041361	0.041851	3	NAT13	0.001144	0.037602	0.040429	3
H1FX	0.00036	0.037602	0.040429	3	NCOA4	0.000806	0.039103	0.044148	3
HADHB	0.008141	0.037602	0.043404	3	NES	0.007265	0.037602	0.044148	3
HDLBP	0.003556	0.046309	0.047607	1	NFIB	0.00204	0.042206	0.053834	1
HIST1H1C	0.009918	0.040504	0.05164	1	NIPBL	0.004622	0.046309	0.052133	1
HIST1H2B	0.000721	0.037602	0.040429	3	NONO	0.00895	0.042895	0.040445	3
HN1	0.004378	0.037602	0.040429	3	NR2F2	0.00171	0.037602	0.040429	3
HNRNPU	0.008161	0.037602	0.040429	3	NUP88b	0.002406	0.042333	0.044148	3
HNRPH1	0.0011	0.048531	0.047607	1	OGN	0.000806	0.041361	0.045032	3
HOMER2	0.010124	0.042333	0.044148	3	OPTN	0.002773	0.04879	0.057663	1
HOXB7	0.002263	0.037602	0.040429	3	ORMDL1	0.00036	0.037602	0.041428	3
HSPA8	0.002599	0.037602	0.047607	1	PABPC3	0.001435	0.037602	0.044148	3
HTRA3	0.000468	0.046309	0.053888	1	PCDH19	0.001003	0.044315	0.05409	1
IFT172	0.00118	0.042751	0.048727	1	PCNA	0.000468	0.044596	0.044238	1
IMPDH2	0.00396	0.040504	0.045032	3	PDHB	0.006271	0.041361	0.052317	1
IRF2BP2	0.001189	0.037602	0.040429	3	PGRMC2	0.000946	0.046309	0.050865	1
JARID1A	0.00666	0.04319	0.048589	1	PKIA	0.000787	0.037602	0.040429	3
JARID1B	0.000465	0.039025	0.040429	3	PLCD1	0.005183	0.037602	0.040429	3
JMJD2B	0.042054	0.037602	0.040429	2	PMPCB	0.000806	0.037602	0.042641	3
JUN	0.001699	0.037602	0.040429	3	PPA2	0.011623	0.039103	0.047607	3
KIAA0101	0.000424	0.047099	0.047607	1	PPAP2B	0.010559	0.041847	0.040429	3
KIAA0368	0.003621	0.037602	0.048589	1	PPP2R5A	0.001915	0.037602	0.043734	3
KIAA1128	0.006271	0.039103	0.041369	3	PRPH	0.04213	0.04014	0.043404	2
KIRREL	0.001618	0.044315	0.048589	1	PRR14	0.0011	0.044161	0.048589	1
KLF3	0.00677	0.048218	0.05409	1	PSMA1	0.002136	0.044596	0.049025	1
KLF5	0.002909	0.037602	0.040429	3	PSMD12	0.002012	0.039025	0.048589	1
KLHDC3	0.009229	0.044315	0.041851	3	PSMD4	0.003401	0.041361	0.048589	1
LAMA3	0.038604	0.037602	0.047429	2	PSMD5	0.00209	0.037602	0.045032	3
LAMB1	0.010723	0.037602	0.047429	3	PTPRM	0.000814	0.037602	0.040429	3
LAMB2	0.00036	0.037602	0.047607	1	PXDN	0.0011	0.037602	0.040429	3
LMO4	0.002249	0.047099	0.041863	1	RAB14	0.006174	0.046605	0.052133	1
LTBP1	0.005152	0.037602	0.040429	3	RAN	0.001801	0.048218	0.054802	1
MAN1A2	0.038487	0.04241	0.047607	2	RANBP5	0.007889	0.049801	0.052781	1
MAP7D1	0.00118	0.050689	0.051925	1	RFC3	0.0011	0.041361	0.043448	3
MAT2A	0.000787	0.037602	0.040429	3	RP9	0.002834	0.044742	0.053014	1
MCM6	0.002643	0.044596	0.054109	1	RPP30	0.001535	0.04879	0.051031	1
MGA	0.004803	0.040504	0.050959	1	RRBP1	0.007963	0.037602	0.040429	3
MIS12	0.00144	0.037602	0.041863	3	RSBN1	0.001635	0.037602	0.040429	3
MLL	0.003762	0.04241	0.048589	1	RUSC2	0.002834	0.037602	0.04716	1
MOCS2	0.002064	0.040504	0.040445	3	SEC31A	0.002617	0.042751	0.048589	1
MRPS10	0.003114	0.039103	0.050312	1	SERPING1	0.00182	0.037602	0.045013	3
MSTN	0.010352	0.037602	0.040429	3	SF3A3	0.004562	0.039124	0.044291	3
MT-ND1	0.005126	0.037602	0.05186	1	SFRS1	0.001271	0.037602	0.041851	3
MT-ND4	0.00182	0.037602	0.040429	3	SFRS3	0.002725	0.041361	0.0522	1
MT-ND5	0.002823	0.037602	0.040429	3	SFRS5	0.003904	0.039103	0.041851	3
MYLK	0.040248	0.037602	0.040429	2	SKP2	0.000468	0.037602	0.040429	3

FDR ≤0.05: Genes differentially expressed.

Table 4 cont.

NAME	FRD L	FDR RS	FDR LB	C	_NAME_	FRD L	FDR RS	FDR LB	C
SLC15A2	0.00204	0.039124	0.050594	1	TMEM34	0.003027	0.044237	0.049224	1
SLC16A3	0.01021	0.037602	0.040429	3	TRAM1	0.0011	0.039556	0.040429	3
SLC25A24	0.002468	0.044315	0.04753	1	TRIP12	0.005508	0.044596	0.044238	1
SLC2A13	0.002981	0.039103	0.047607	1	TSNAXIP1	0.001984	0.037602	0.040429	3
SLC35A5	0.001219	0.046605	0.057663	1	TSPAN7	0.003952	0.049985	0.051643	1
SLC44A2	0.001865	0.037602	0.040429	3	TUBA1A	0.0011	0.037602	0.041863	3
SLIT2	0.006255	0.037602	0.040429	3	TUBA1C	0.001496	0.044596	0.047607	1
SLMO2	0.001852	0.048986	0.048589	1	TUBA4A	0.0011	0.048525	0.048589	1
SMAD9	0.002879	0.037602	0.040429	3	TUBA4A	0.0011	0.037602	0.040429	3
SMARCA5	0.0011	0.042333	0.041863	3	TUBB6	0.01307	0.037602	0.044148	3
SNAP91	0.011051	0.037602	0.043448	3	TXNDC1	0.001039	0.046605	0.047429	1
SOX4	0.000169	0.037602	0.040429	3	UBE2D3	0.001825	0.039103	0.045032	3
SPG20	0.00148	0.037602	0.040429	3	UBE3A	0.003292	0.037602	0.040429	3
SPG21	0.005621	0.047063	0.053834	1	UBLCP1	0.001706	0.046605	0.057663	1
SRPRB	0.003174	0.044315	0.045032	1	UGP2	0.006539	0.044237	0.048727	1
STAT1	0.001538	0.039556	0.043734	3	UMPS	0.002545	0.042751	0.048727	1
STCH	6.89E-05	0.037602	0.040429	3	UQCRFS1	0.001801	0.045537	0.053834	1
SUV39H2	0.000205	0.04319	0.048589	1	VPRBP	0.00951	0.04879	0.053888	1
SYPL1	0.002061	0.048525	0.053834	1	WDR54	0.003037	0.037602	0.041851	3
TAF13	0.00077	0.037602	0.040429	3	WNK2	0.000732	0.047099	0.054948	1
TBC1D23	0.000397	0.039025	0.041851	3	WWC3	0.003333	0.037602	0.040429	3
TBL1X	0.00148	0.037602	0.040429	3	XBP1	0.011264	0.037602	0.043404	3
TIMM8A	0.00118	0.039103	0.041863	3	ZBTB16	0.026285	0.037602	0.043404	24
TLOC1	0.008318	0.042333	0.050988	1	ZFP106	0.002269	0.037602	0.043404	3
TM9SF2	0.001008	0.047099	0.048589	1	ZNF207	0.005523	0.041847	0.050312	1
TMED10	0.001845	0.037602	0.040429	3	ZNF22	0.001902	0.037602	0.041851	3
TMED2	0.001869	0.039949	0.044291	3	ZNF503	6.89E-05	0.037602	0.040429	3
TMEM132C	0.046071	0.04241	0.044238	2	-				

FDR ≤0.05: Genes differentially expressed.

For the contrast comparing both ages in the Piau breed (P40vsP70), the clustering analysis suggested a separation of the whole data in 400 different clusters. This conclusion was taken based on a graph, according to the maximum curvature of the proposed clusters distributed in terms of numbers of genes in each (not shown). Among these 400 clusters, the tenth (10^{o}) cluster proposed presented the lowest average of FDR values in each of the three protocol analysis (Loess normalization method with no background correction, Robust Spline and Loess normalization method background corrected): 0.018202, 0.030494 and 0.02914, respectively. The number of oligos with HGNC information included in this cluster was equal to 182. When proceeding another clustering analysis within this list of 182 genes, 30

sub clusters were suggested, but six of them (1st, 2nd, 3rd, 4th, 6th and 9th), presented the lowest FDR averages among the three analysis (Table 5 and 6). Again, the FDR values presented in this contrast are more variable between both analyses (Robust Spline and Loess).

Table 5. Six sub clusters more informative within the 50 suggested for the contrast between ages in the Piau breed in terms of FDR values obtained after analysis. Number of the cluster, number of genes in the cluster (FREQ) and FDR average of the cluster respectively in the previous analysis (L) and current analysis background corrected before using the Robust Spline (RS) and Loess (LB) as normalization methods.

CLUSTER	_FREQ_	FRD L	FDR RS	FDR LB
1	17	0.016006	0.046569	0.04714
2	26	0.006318	0.032156	0.022784
3	41	0.006247	0.009404	0.010385
4	13	0.005577	0.048109	0.031113
6	5	0.027839	0.012425	0.010327
9	18	0.003205	0.018007	0.022081
	120			

Table 6. Name of the gene, respective FDR values and number of the sub cluster (C) more informative within the 30 suggested for the contrast between ages in the Piau breed in terms of FDR values obtained after analysis.

NAME	FRD L	FDR RS	FDR LB	C	_NAME_	FRD L	FDR RS	FDR LB	C
ABRA	0.005982	0.051588	0.047704	1	FGL2	8.5E-05	0.029729	0.013451	2
ACTA1	0.008056	0.030074	0.022149	2	FOXP1	0.005242	0.03482	0.024795	2
ACTN4	0.002178	0.029131	0.017673	2	FSD2	0.000633	0.013403	0.011602	3
ALDH2	0.009319	0.007181	0.013451	3	FXYD1	0.001459	0.020517	0.026522	9
ALDOA	0.010809	0.053166	0.031442	4	FYCO1	3.68E-05	0.016078	0.028116	9
AMPD1	0.009319	0.041168	0.027264	4	GADD45G	0.001891	0.020448	0.023358	9
AMPD1	0.014406	0.0314	0.019484	2	GJB1	0.008785	0.048813	0.049605	1
ANXA2	0.019806	0.008809	0.013278	3	GPR37	0.001891	0.030074	0.014573	2
APOBEC2	0.011705	0.035452	0.019422	2	GSTP1	0.015858	0.047778	0.04423	1
ART3	0.000309	0.028618	0.030917	2	H2AFJ	0.021836	0.050043	0.047953	1
ASB2	3.68E-05	0.029131	0.023358	2	HN1	0.012115	0.029577	0.019422	2
AVPI1	0.019806	0.007181	0.010024	3	HRASLS	0.003411	0.025047	0.016545	9
BIN1	0.02247	0.014241	0.009051	6	IFT122	0.000621	0.037525	0.026522	2
BRP44L	0.02154	0.041547	0.049187	1	IGSF1	0.007519	0.040598	0.051631	1
C16orf14	0.02027	0.054641	0.050876	1	ITGA7	0.007326	0.033898	0.019862	2
C1orf93	0.024384	0.010697	0.009051	6	ITGB1BP2	0.016754	0.010755	0.011602	3
C20orf58	0.004226	0.014448	0.027264	9	KBTBD10	0.012062	0.040598	0.018764	2
C3orf45	0.004838	0.035194	0.031442	2	LSAMP	0.000931	0.011036	0.015812	3
C6orf115	0.009319	0.015887	0.022973	9	MACROD1	0.029886	0.015416	0.009051	6
CA3	0.003874	0.00876	0.009051	3	MCM6	0.005898	0.023126	0.029682	9
CAPN3	0.001275	0.031682	0.022149	2	MYBPC1	0.011705	0.029577	0.021642	2
CAV1	0.000755	0.0314	0.027264	2	MYH1	0.001324	0.021078	0.015141	9
CAV3	0.002178	0.007181	0.011602	3	MYL1	0.009319	0.007181	0.005686	3
CD36	0.001416	0.011036	0.009051	3	MYL7	0.013173	0.043763	0.056451	1
CD9	0.003602	0.01175	0.018764	9	MYOM2	0.003321	0.010809	0.027264	9
CITED1	0.023809	0.046465	0.035827	1	MYOZ1	0.000855	0.011036	0.010762	3
CKM	0.009319	0.003131	0.002038	3	NAT13	0.015466	0.012687	0.015812	3
COMT	0.008056	0.040598	0.044535	1	NDUFA9	0.017628	0.047778	0.041195	1
CPM	0.00558	0.012687	0.013451	3	NEB	0.00331	0.009296	0.008055	3
CREM	0.000166	0.021992	0.012465	9	NRAP	0.000166	0.04753	0.038535	4
CSDA	0.006463	0.008354	0.009051	3	NRAP	9.2E-05	0.012687	0.011486	3
CSDA	0.004017	0.021979	0.013451	9	PFKM	0.002198	0.007181	0.013278	3
CTSL1	0.009319	0.054842	0.024579	4	PKM2	0.014416	0.030074	0.024579	2
CXXC5	0.010993	0.044988	0.049136	1	PLCL2	0.000617	0.057637	0.027264	4
DDIT4L	0.006815	0.030074	0.022149	2	POPDC3	0.00081	0.007181	0.00844	3
DHRS7C	0.03196	0.009486	0.01221	6	PPP1R3A	0.003411	0.00876	0.009142	3
DMN	0.008427	0.03348	0.033018	2	PPP1R3C	0.030495	0.012287	0.01227	6
DPP4	0.000633	0.04348	0.032961	4	PRX	0.019806	0.013403	0.017254	3
DUSP26	0.000886	0.043763	0.033996	4	PTN	0.000582	0.007181	0.008966	3
DYSFIP1	0.004277	0.007181	0.005686	3	PTPLA	0.01453	0.007181	0.010671	3
EIF4EBP1	0.011633	0.00876	0.011958	3	RAMP1	0.016194	0.012287	0.008301	3
ENO3	0.001288	0.007181	0.005686	3	RRAGD	0.012283	0.03708	0.018252	2
FABP4	0.014998	0.047121	0.049468	1	RTN4	3.68E-05	0.007181	0.011602	3
FAM134B	0.005805	0.051458	0.038174	4	SCG5	0.009319	0.010307	0.01227	3
FEZ2	6.33E-05	0.012687	0.010024	3	SLC25A4	0.002178	0.008264	0.009051	3

FDR ≤0.05: Genes differentially expressed.

Table 6 cont.

NAME	FRD L	FDR RS	FDR LB	C	_NAME_	FRD L	FDR RS	FDR LB	C
SLC30A7	0.007888	0.031682	0.029814	2	TPM3	0.00965	0.010697	0.013278	3
SLC31A2	0.015858	0.044531	0.041063	1	TSPAN2	9.08E-05	0.015648	0.019862	9
SMTN	0.011231	0.01228	0.015892	3	TTN	0.002228	0.011036	0.005686	3
SNAPC2	0.012855	0.043695	0.032632	4	TUBA1B	0.001459	0.013403	0.009051	3
STAC3	0.004481	0.028618	0.023367	2	TUBA1C	0.003205	0.020102	0.027264	9
STBD1	0.023847	0.047841	0.052181	1	TUBA1C	0.001288	0.018551	0.021642	9
TCAP	0.013736	0.011926	0.009051	3	TUBA3E	0.000633	0.007181	0.012349	3
TEAD4	0.008254	0.021992	0.023358	9	TUBA4A	0.000309	0.007181	0.008055	3
TMEM95	0.007519	0.029577	0.032961	2	TUBA4A	0.00298	0.010755	0.022149	9
TMOD4	0.000116	0.00876	0.014699	3	TXNIP	0.016364	0.043119	0.049136	1
TNFRSF12	0.025586	0.050468	0.041195	1	UBE2G1	0.004592	0.048911	0.027264	4
TNNI2	0.005266	0.00876	0.007889	3	UGP2	0.003205	0.013914	0.021642	9
TOMM20	0.002442	0.037525	0.017673	2	UNC45B	0.013204	0.038538	0.031442	4
TPD52	0.000425	0.048813	0.027473	4	USP13	0.005384	0.030074	0.017673	2
TPM1	3.68E-05	0.007181	0.005686	3	ZFP106	0.003874	0.052414	0.031442	4

FDR ≤0.05: Genes differentially expressed.

The contrast comparing different genetic groups, although presenting more genes DE (which is expected), similarities of FDR values among the three protocol analyses tested were more clearly seen. Otherwise, it is suggested that the differential effects of normalization method applied were less sensitive when comparing stages of development within one breed; in that case, the Piau.

4.3. Gene Ontology Analysis

The GO analysis was applied to complement in terms of gene functionality the comparisons between differences and similarities among the results of the protocol analyses tested. The p-values proposed for each term in each analysis is in accordance with a modified Fisher exact test, based on a EASE score. It is used to determine whether the proportion of those genes falling into each GO category differs by group (gene-enrichment in annotation terms). The smaller, the more enriched.

Firstly, the lists of genes which were indicated as DE in each of the two contrast according to the current analyses testing two normalization method (n=34 for C40vsP40 and n=37 for P40 vs P70), but not

in the previous one, were analyzed on Gene Ontology to check in which GO term they could be included (Table 7a, 7b, respectively).

Table 7a- GO category: Biological Processes (BP), Cellular Component (CC) or Molecular Function, GO term, number of genes evolved and p-value information according to the GO analysis of the contrast between both genetic groups at 40 d of gestation under Robust Spline (RS*) or Loess (L*) normalization methods.

Biological Processes Term	Count	PValue	Genes
microtubule-based process	4	0.009787	TUBB2A, KIF5A, ESPL1, KIF13B
microtubule-based movement	3	0.016932	TUBB2A, KIF5A, KIF13B
negative regulation of catalytic activity	3	0.085864	SH3BP5, CST3, FABP4
transcription, DNA-dependent	3	0.093931	HIF3A, CRCP, MED14
RNA biosynthetic process	3	0.096119	HIF3A, CRCP, MED14
Cellular Component Term			
microtubule cytoskeleton	4	0.073807	TUBB2A, KIF5A, ESPL1, KIF13B
microtubule	3	0.086272	TUBB2A, KIF5A, KIF13B
Molecular Function Term			
enzyme inhibitor activity	4	0.013034	SH3BP5, CST3, PCSK1N, PPP1R14A

Table 7b- GO category: Biological Processes (BP), Cellular Component (CC) or Molecular Function, GO term, number of genes evolved and p-value information according to the GO analysis of the contrast between ages in the Piau breed under Robust Spline (RS*) or Loess (L*) normalization methods.

Biological Processes Term	Count	PValue	Genes
tissue morphogenesis	6	2.07E-05	TNNT2, ACTC1, MACF1, MYL2, LMO4, FST
myofibril assembly	3	7.29E-04	TNNT2, ACTC1, MYL2
muscle tissue morphogenesis	3	0.00121	TNNT2, ACTC1, MYL2
cardiac muscle tissue morphogenesis	3	0.00121	TNNT2, ACTC1, MYL2
actomyosin structure organization	3	0.001302	TNNT2, ACTC1, MYL2
cellular component assembly involved in morphogenesis	3	0.00215	TNNT2, ACTC1, MYL2
striated muscle cell development	3	0.00444	TNNT2, ACTC1, MYL2
cytoskeleton-dependent intracellular transport	3	0.00444	TNNT2, ACTC1, MACF1
muscle cell development	3	0.005132	TNNT2, ACTC1, MYL2
cardiac muscle tissue development	3	0.005313	TNNT2, ACTC1, MYL2
muscle organ development	4	0.007466	TNNT2, ACTC1, MYL2, FHL1
heart morphogenesis	3	0.008585	TNNT2, ACTC1, MYL2
actin cytoskeleton organization	4	0.00901	TNNT2, ACTC1, MYL2, PLS3
cytoskeleton organization	5	0.009065	TNNT2, ACTC1, MACF1, MYL2, PLS3
actin filament-based process	4	0.010729	TNNT2, ACTC1, MYL2, PLS3
cardiac myofibril assembly	2	0.011478	ACTC1, MYL2
striated muscle cell differentiation	3	0.012286	TNNT2, ACTC1, MYL2
cardiac cell development	2	0.020947	ACTC1, MYL2
cardiac muscle cell development	2	0.020947	ACTC1, MYL2
striated muscle tissue development	3	0.02173	TNNT2, ACTC1, MYL2
muscle cell differentiation	3	0.022418	TNNT2, ACTC1, MYL2
muscle tissue development	3	0.023819	TNNT2, ACTC1, MYL2
cellular component morphogenesis	4	0.039576	TNNT2, ACTC1, MACF1, MYL2
ventricular cardiac muscle morphogenesis	2	0.041472	TNNT2, MYL2
heart contraction	2	0.043317	ACTC1, MYL2
cardiac muscle cell differentiation	2	0.043317	ACTC1, MYL2
actin filament-based movement	2	0.043317	TNNT2, ACTC1
heart process	2	0.043317	ACTC1, MYL2
cardiac cell differentiation	2	0.052492	ACTC1, MYL2
heart development	3	0.063671	TNNT2, ACTC1, MYL2

Table 7b cont.

Cellular Component Term	Count	PValue	Genes
sarcomere	4	8.89E-04	TNNT2, ACTC1, MYL2, HSPB1
myofibril	4	0.001275	TNNT2, ACTC1, MYL2, HSPB1
contractile fiber part	4	0.001343	TNNT2, ACTC1, MYL2, HSPB1
contractile fiber	4	0.001635	TNNT2, ACTC1, MYL2, HSPB1
actin cytoskeleton	4	0.015056	TNNT2, ACTC1, MACF1, MYL2
Molecular Function Term			
cytoskeletal protein binding	5	0.009408	TNNT2, ACTC1, MACF1, MYL2, PLS3
calcium ion binding	6	0.016828	CDH15, MACF1, MYL2, PLS3, S100A13, CDH11
actin binding	4	0.01696	TNNT2, MACF1, MYL2, PLS3
myosin binding	2	0.036653	ACTC1, MYL2

Even though not all the 34 genes respective to the contrast between both genetic groups at 40 d of gestation were recognized by Gene Ontology, interesting genes appeared to be involved with important mechanisms acting on muscle development. The Table 7a represents genes like SH3BP5, CST3, PCSK1N and PPP1R14A related to microtubule-based process and TUBB2A, KIF5A, ESPL1, KIF13B, related to enzyme inhibitor activity. These genes composed the two significant GO terms ($p \leq 0.05$) in this contrast and interestingly, they were all considered as differentially expressed (FDR≤0.05) exclusively in the analysis with Robust Spline normalization method.

The table 7b represents few genes (9 out of 37) recognized by the Gene Ontology evolved with potential mechanisms, specially related to biological processes, taking place in earlier or later stages of the myogenesis process when comparing both prenatal ages in the Piau breed. The genes HSPB1, PLS3 and S100A13 were exclusively DE in the analysis with Robust Spline normalization method, while ACTC1 and CDH15 were exclusively DE in the analysis with Loess normalization method in this contrast.

As a second strategy, we preceded the GO analysis with the list of genes obtained after clustering analysis. For that, each list of genes (n= 237 and n=120) suggested to joint similar genes in terms of FDR values among the three different protocol analyses tested in each of the contrasts analyzed (C40vsP40 and P40vsP70, respectively), was analyzed according to the Gene Ontology classifications.

The figure 2 represents the proportional differences of "GO-terms" in each category (Biological Process, Cellular Component and Molecular Function) for each of these two contrasts. The distribution of each term in each contrast was proportionally similar, even though the first contrast presented biggest numbers, once it had more genes DE comparing with the second one.

The table 8 (a-c) specifies terms found significant ($p < 0.05$) with the respective number of genes and the *p value* proposed by the GO analysis for the three categories.

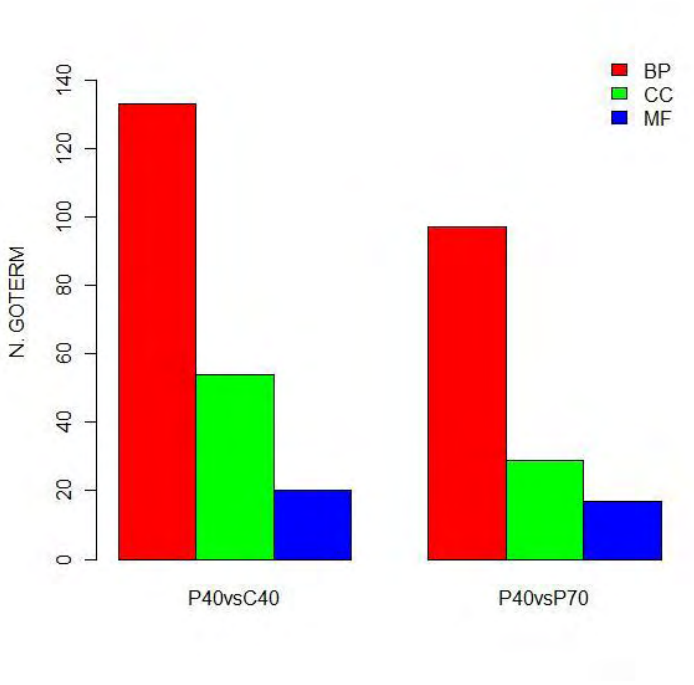


Figure2 – Graph showing the proportion of GO terms included in each category according to the list proposed by the clustering analysis for each contrast. BP- Biological Processes category; CC- Cellular Component category; MF- Molecular Function category.

Table8a. Over-represented gene ontology Biological Processes for genes with similar FDR values among the three protocol analysis according to the clustering analysis of each contrast.

	C40 vs P40		P40 vs P70	
Biological Process Gene Ontology Term	Count	PValue	Count	PValue
response to mechanical stimulus	5	5.21E-03	4	6.12E-03
response to endogenous stimulus	13	4.90E-03	8	1.82E-02
response to hormone stimulus	12	6.41E-03	8	1.11E-02
response to inorganic substance	9	4.23E-03	5	4.78E-02
cellular macromolecular complex subunit organization	13	1.76E-03	8	9.65E-03
cellular macromolecular complex assembly	12	2.18E-03	8	5.24E-03
response to drug	9	5.78E-03	6	1.46E-02
macromolecular complex subunit organization	17	1.62E-02	10	4.62E-02
response to steroid hormone stimulus	7	3.36E-02	6	9.08E-03
protein polymerization	4	2.32E-02	5	3.09E-04
macromolecular complex assembly	16	1.95E-02	10	3.25E-02
chromatin remodeling	8	5.81E-06		
response to cAMP	7	1.26E-05		
response to reactive oxygen species	8	4.08E-05		
response to hydrogen peroxide	7	6.73E-05		
transmembrane receptor protein serine/threonine kinase signaling pathway	8	3.04E-04		
response to corticosteroid stimulus	7	6.73E-04		
positive regulation of macromolecule metabolic process	22	2.33E-03		
response to glucocorticoid stimulus	6	2.90E-03		
response to oxidative stress	8	4.53E-03		
positive regulation of transcription	16	4.74E-03		
positive regulation of gene expression	16	6.19E-03		
response to extracellular stimulus	9	6.44E-03		
transforming growth factor beta receptor signaling pathway	5	6.66E-03		
regulation of cell adhesion	7	7.45E-03		
proteasomal ubiquitin-dependent protein catabolic process	6	9.03E-03		
proteasomal protein catabolic process	6	9.03E-03		
androgen receptor signaling pathway	4	1.01E-02		
biopolymer methylation	5	1.08E-02		
positive regulation of transcription, DNA-dependent	13	1.66E-02		
positive regulation of macromolecule biosynthetic process	16	1.70E-02		
BMP signaling pathway	4	1.74E-02		
positive regulation of RNA metabolic process	13	1.76E-02		
chromosome organization	13	1.87E-02		
chromatin organization	11	2.04E-02		
chromatin modification	9	2.18E-02		
regulation of cell cycle	10	2.33E-02		
positive regulation of cellular biosynthetic process	16	2.46E-02		
DNA metabolic process	13	2.50E-02		

Table 8a cont.

	C40 vs P40		P40 vs P70	
Biological Process Gene Ontology Term	Count	PValue	Count	PValue
positive regulation of biosynthetic process	16	2.76E-02		
anterior/posterior pattern formation	6	3.11E-02		
ribonucleoside monophosphate biosynthetic process	3	3.31E-02		
DNA methylation	3	3.58E-02		
response to acid	3	3.58E-02		
DNA alkylation	3	3.58E-02		
steroid hormone receptor signaling pathway	4	3.58E-02		
response to organic substance	16	3.64E-02		
regionalization	7	3.73E-02		
ribonucleoside monophosphate metabolic process	3	3.86E-02		
response to abiotic stimulus	10	4.16E-02		
cellular response to extracellular stimulus	4	4.59E-02		
negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle	4	4.77E-02		
Anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	4	4.77E-02		
chromatin-mediated maintenance of transcription	2	4.90E-02		
blood vessel morphogenesis	7	4.92E-02		
muscle contraction			16	6.98E-14
muscle system process			16	2.80E-13
muscle organ development			15	1.04E-10
striated muscle contraction			8	1.97E-08
cellular component assembly involved in morphogenesis			7	1.22E-07
myofibril assembly			6	2.18E-07
actomyosin structure organization			6	1.01E-06
muscle cell differentiation			9	1.36E-06
muscle cell development			7	1.82E-06
striated muscle cell development			6	2.37E-05
striated muscle cell differentiation			7	2.56E-05
actin filament-based process			10	3.13E-05
sarcomere organization			4	6.00E-05
glucose metabolic process			8	7.03E-05
hexose metabolic process			8	2.88E-04
muscle thin filament assembly			3	4.32E-04
skeletal myofibril assembly			3	6.45E-04
monosaccharide metabolic process			8	6.87E-04
cellular protein complex assembly			7	7.30E-04
actin cytoskeleton organization			8	7.63E-04
striated muscle tissue development			6	1.17E-03

Table 8a cont.

	C40 vs P40		P40 vs P70	
Biological Process Gene Ontology Term	Count	PValue	Count	PValue
actin filament organization			5	1.33E-03
muscle tissue development			6	1.46E-03
alcohol catabolic process			5	2.06E-03
actin cytoskeleton organization			8	7.63E-04
striated muscle tissue development			6	1.17E-03
actin filament organization			5	1.33E-03
muscle tissue development			6	1.46E-03
alcohol catabolic process			5	2.06E-03
cytoskeleton organization			10	2.43E-03
glycolysis			4	3.74E-03
muscle maintenance			3	3.78E-03
generation of precursor metabolites and energy			8	4.82E-03
cellular component morphogenesis			9	4.88E-03
response to calcium ion			4	5.82E-03
cardiac muscle tissue development			4	6.43E-03
glucose catabolic process			4	6.75E-03
cardiac muscle contraction			3	6.95E-03
response to progesterone stimulus			3	7.69E-03
heart process			3	1.01E-02
heart contraction			3	1.01E-02
actin filament-based movement			3	1.01E-02
hexose catabolic process			4	1.09E-02
monosaccharide catabolic process			4	1.18E-02
regulation of muscle contraction			4	1.22E-02
muscle tissue morphogenesis			3	1.38E-02
cardiac muscle tissue morphogenesis			3	1.38E-02
regulation of cellular component biogenesis			5	1.48E-02
cellular carbohydrate catabolic process			4	1.90E-02
protein complex assembly			9	1.91E-02
protein complex biogenesis			9	1.91E-02
purine ribonucleotide transport			2	1.98E-02
muscle thick filament assembly			2	1.98E-02
purine nucleotide transport			2	1.98E-02
myosin filament assembly or disassembly			2	1.98E-02
myosin filament assembly			2	1.98E-02
negative regulation of protein kinase activity			4	2.02E-02
negative regulation of kinase activity			4	2.21E-02
muscle fiber development			3	2.26E-02

Table 8a cont.

	C40 vs P40		P40 vs P70	
Biological Process Gene Ontology Term	Count	PValue	Count	PValue
negative regulation of transferase activity			4	2.61E-02
nucleotide transport			2	2.63E-02
sterol homeostasis			3	2.64E-02
cholesterol homeostasis			3	2.64E-02
fructose 1,6-bisphosphate metabolic process			2	3.28E-02
cardiac muscle fiber development			2	3.28E-02
positive regulation of cytoskeleton organization			3	3.60E-02
carbohydrate catabolic process			4	3.61E-02
microtubule-based movement			4	3.95E-02
lipid homeostasis			3	4.52E-02
membrane raft organization			2	4.57E-02
cytoskeleton-dependent intracellular transport			3	4.68E-02

Table 8b. Over-represented gene ontology Cellular Component for genes with FDR values similar among the three protocol analysis according to the clustering analysis.

	C40 vs P40		P40 vs P70	
Cellular Component Gene Ontology Term	Count	PValue	Count	PValue
non-membrane-bounded organelle	49	2.21E-03	26	2.17E-02
intracellular non-membrane-bounded organelle	49	2.21E-03	26	2.17E-02
proteasome complex	7	1.10E-04		
extracellular matrix part	8	6.62E-04		
proteinaceous extracellular matrix	13	6.95E-04		
organelle membrane	27	1.10E-03		
extracellular matrix	13	1.33E-03		
chromosome	15	1.91E-03		
cytosol	30	2.05E-03		
membrane-enclosed lumen	38	2.32E-03		
chromosomal part	13	3.37E-03		
chromatin	9	3.70E-03		
endomembrane system	20	4.15E-03		
nuclear chromatin	5	4.91E-03		
laminin complex	3	5.29E-03		
organelle lumen	36	5.59E-03		
intracellular organelle lumen	35	6.99E-03		
extracellular region part	22	8.55E-03		
cytoplasmic membrane-bounded vesicle	15	9.17E-03		
collagen	4	9.39E-03		
fibrillar collagen	3	9.46E-03		
Golgi apparatus part	10	1.17E-02		
membrane-bounded vesicle	15	1.19E-02		
chromatin remodeling complex	5	1.20E-02		
nuclear lumen	29	1.28E-02		
coated vesicle	7	1.49E-02		
basement membrane	5	1.65E-02		
basal lamina	3	1.87E-02		

Table 8b cont.

	C40 vs P40		P40 vs P70	
Cellular Component Gene Ontology Term	Count	PValue	Count	PValue
melanosome	5	2.54E-02		
pigment granule	5	2.54E-02		
cytoplasmic vesicle	15	3.06E-02		
mitochondrial envelope	11	3.79E-02		
vesicle	15	4.15E-02		
coated membrane	4	4.43E-02		
membrane coat	4	4.43E-02		
organelle envelope	14	4.75E-02		
envelope	14	4.85E-02		
transport vesicle	4	4.97E-02		
myofibril			15	9.60E-15
contractile fiber			15	3.29E-14
sarcomere			14	4.74E-14
contractile fiber part			14	3.14E-13
actin cytoskeleton			17	1.39E-11
striated muscle thin filament			6	2.92E-08
I band			6	2.76E-05
cytoskeleton			23	4.83E-05
Z disc			5	2.19E-04
myosin complex			5	8.28E-04
pseudopodium			3	8.57E-04
soluble fraction			9	9.47E-04
cytoskeletal part			16	1.09E-03
A band			3	4.71E-03
myosin filament			3	6.63E-03
stress fiber			3	1.05E-02
actin filament bundle			3	1.22E-02
actomyosin			3	1.32E-02
muscle thin filament tropomyosin			2	2.57E-02
cell-substrate adherens junction			4	3.16E-02
cell-substrate junction			4	3.63E-02
cell fraction			13	4.63E-02

Table 8c. Over-represented gene ontology Molecular Function for genes with similar FDR values among the three protocol analysis according to the clustering analysis.

	C40 vs P40		P40 vs P70	
Molecular Function Gene Ontology Term	Count	PValue	Count	PValue
structural molecule activity	15	1.58E-02	16	7.35E-06
GTPase activity	8	1.17E-02		
extracellular matrix structural constituent	5	1.79E-02		
protein C-terminus binding	6	2.44E-02		
double-stranded DNA binding	5	2.65E-02		
structure-specific DNA binding	6	2.71E-02		
promoter binding	4	2.86E-02		
GTP binding	10	2.98E-02		
chromatin binding	6	3.08E-02		
solute:hydrogen symporter activity	3	3.38E-02		
growth factor binding	5	3.41E-02		
guanyl ribonucleotide binding	10	3.45E-02		
guanyl nucleotide binding	10	3.45E-02		
transcription regulator activity	26	4.37E-02		
structural constituent of muscle			8	6.39E-09
cytoskeletal protein binding			17	6.84E-08
actin binding			13	7.68E-07
identical protein binding			12	2.00E-03
fructose binding			2	3.69E-02
protein heterodimerization activity			5	4.10E-02
titin binding			2	4.29E-02
alpha-actinin binding			2	4.29E-02

As we can see in the Table 8a, in agreement between both contrasts, Biological Process terms such as, response to endogenous stimulus, response to hormone stimulus, cellular macromolecular complex subunit organization, macromolecular complex subunit organization and macromolecular complex assembly are more relevant in terms of number of genes related with. Otherwise, cytoskeleton organization, actin filament-based process and muscle contraction are one of the most relevant terms specifically related to the P40 vs P70 contrast.

As common terms of the Cellular Component category (Table 8b) in both contrasts, non-membrane-bounded organelle and intracellular non-membrane-bounded organelle were the only ones, but

the terms suggested only in the second contrast were related to muscle development and structure: myofibril, contractile fiber and cytoskeleton, for example.

Similarly, the structural molecule activity term was the only one in agreement between both contrasts for the Molecular Function category (Table 8c), but the others related to the second contrast were also closely related to muscle development.

5. Discussion

Our study has shown that the different protocols of microarray analysis including background correction procedure, or not, and two types of normalization methods, differ markedly in terms of number of genes detected as differentially expressed and precision.

Although with a relevant number of normalization methods frequently being proposed, it has been difficult to decide which method performs better than the others. Some papers (e.g. Park et al., 2003), concluded that the intensity dependent normalization method (most commonly applied) performs better than traditional methods, like the simpler global normalization method in many cases of microarray experiments. In our case, although studying only one data set with few slides, our findings provided some guidance on the selection of normalization methods. As shown in the figure2 (Appendix 3), scatherplots helps exploratory analysis for diagnoses of data distribution, characterizing heteroscedasticity and dependences. Also, allows us to see differences in the between array variations and intensity dependent trends when comparing normalization methods (Bolstad et al., 2003). Both methods reduced the variance at all intensity levels in comparison to the non normalized graph, but the Robust Spline performed slightly better for this dataset.

However, a trade-off among the results was observed when comparing the current protocol analyses tested after statistical analyses. The Robust Spline normalization method was more accurate to detect genes differentially expressed in the contrast between both genetic groups at 40d of gestation, while the contrast between both ages in the Piau breed presented genes with significant FDR values more variable between both methods. Also, the clustering analysis for this last contrast presented more number

of sub clusters attempting to joint genes closely related in terms of FDR values, suggesting that breed-specific genes are less sensitive to different normalization (and protocols) methods applications.

Robust Spline normalization method has been adopted by Illumina Whole-Genome Expression BeadChips (Du et al., 2008; Shi et al., 2009), and Degenkolbe et al., (2009) affirmed that the Robust Spline method for within array normalization joined with the quantile method for between array normalization, yield the smallest differences between arrays with respect to the position of the median, the variation and the shape of the distribution curve between arrays. According to Boldstad et al., (2003), the quantile method has performed favorably, both in terms of speed and bias criteria, and therefore should be used in preference to other methods. The same authors concluded that when making pairwise comparisons, the quantile method gave the smallest distance between arrays and these distances also remained fairly constant across intensities.

According to Soler et al., (2004), splines smoothing methods give better adjustment than methods currently used for microarray normalization procedure, like the global and Loess adjustments - controlling the dye bias and allowing a clearer identification of differently expressed genes in these experiments. Even though being one of the most recognized normalization method for two-color microarray experiments, the Loess normalization method proposed by Yang et al. (2002), failed to be the best choice when compared with others non-linear based approaches (Workman et al., 2002).

To assess and discuss the data, the clustering analysis based on FDR values obtained by each type of protocol analysis also helped. It was clear that the FDR values obtained according to the previous protocol with the Loess normalization method were constantly lower comparing with the other results obtained by both current protocols for most of the genes. Because two different FDR methods were used between the previous (*q value*, Storey and Tibshiran, 2003) and the current both analyses (BH method, Benjamini and Hochberg, 1995), further speculation are limited, but still, the clustering analysis pointed out genes potentially more similar in terms of FDR values, especially in a breed-specific manner.

Among many different methods developed to assess the false discovery rate (FDR), Benjamini and Hochberg (1995) can be considered as one of the most powerful procedures (Reiner et al., 2003) with

known operating characteristics. Similarly, Storey and Tibshiran (2003) proposed a *qvalue* as an extension of this method to detect false discovery rate (FDR). It proposes to estimate a measure of significance for each feature that meets the practical goals of the genome-wide study and that is easily interpreted in terms of the simultaneous testing of thousands of features. Ponsuksili et al., (2007) adopted this procedure for controlling FDR. However, a wide numbers of FDR methods have been applied in different studies: Benjamini et al., (2001) (Lobjois et al., 2008); Benjamini & Hochberg (1995) (Liu et al., 2009); Holm (1979) (Zhao et al., 2003) and McLachlan et al., (2006) (Lehnert et al., 2007).

According to Storey and Tibshirani (2003), the original BH FDR methodology is too conservative for genomics applications. They assure that the number of gene differentially expressed in a specific experiment above this method was almost half least small than the number detected by their *q-value* method. However, Aubert et al., (2004) conclude that the *q-value* gives a too optimistic view of the probability for the gene to be false positive. They generally found *q-values* lower than others traditional FDR methods (like a local one proposed by them) because it is computed using all the genes that are more significant than a gene *i*. Obviously a gene whose *p-value* is near to the threshold *t* does not have the same probability to be differentially expressed than a gene whose *p-value* is close to zero.

Even though the whole context of the statistical method applied to find the *p-values* was different between the previous and the two new current analyses, we could also suggest that the FDR procedure applied in our previous analysis (Storey and Tibshirani, 2003) was too flexible in the process of jugging genes as differentially expressed. Curiously, the current analyses detected just a few or even no genes differentially expressed (FDR≤0.05) in two (C40vsC70 and C70vsP70, respectively) out of the four contrasts using the BH FDR method.

As pointed out by Reiner et al., (2003), gene expression measurement errors may be dependent due to factors related to the RNA source, the normalization process and the pooled variability estimation. Multiple testing of such data will produce correlated test statistics. Thus, it is essential to account for the dependency structure between the test statistics. Measurement error of microarray data tends to be positively dependent, and simple FDR controlling procedures such as the BH copes with such

dependency. In this way, the empirical Bayes approach (Smyth, 2004) also agrees with this statement, once it is equivalent to shrinkage of the estimated sample variances towards a pooled estimate, resulting in far more stable inference when the number of arrays is small. Moreover, the empirical Bayes moderated *t-test* is one of the most popular algorithms for microarray differential expression which have the characteristic of "borrowing" information between genes (Ritchie et al., 2007). This is especially interesting when the degree of freedom available to estimate gene-wise standard deviation are too small, like in cases there are just two or few slides for direct comparisons.

Proposing to account the unbalance existing in the loop connected design the previous analysis based on mixed model (Tempelman, 2008) resulted in extended lists of genes DE in all four contrasts under question comparing with the current analyses. Xu & Faisal (2010) concluded that the factorial analysis applied on their microarray data properly estimated the potential effects of an experimental condition that was not directly included in the hybridization, but exposed simultaneously to a multitude of environmental factors (four) that are in continuous fluctuation. The current microarray experiment, considering two factors, was firstly designed in a connected loop aiming to attempt statistical power in an economic way. However, as we could test here, statistical analyses considering direct comparisons separately accounting for treatments and factors in independent experiments is also possible, without falling back on a common reference design extensively proved as statistically inefficient (e.g., Vinciotti et al., 2005).

Although the context of this study is behind to the functionality of genes, the Gene Ontology analysis, in general, showed that genes DE clustered among the three protocol analyses were functionally designated to the proper contrast in context (C40vsP40 or P40vsP70). Moreover, in somehow, this type of analysis strengthened the idea that the Robust Spline method were able to identify more genes differentially expressed (FDR ≤ 0.05) related with important GO terms involved with the muscle development process, even considering that only few genes found DE exclusively by the two current protocol analyses were recognized by the Gene Ontology. Clustering gene expression data groups together genes of known similar function (Eisen et al., 1998). The same authors proposed that

coexpressed genes of known function poorly characterized, or novel genes, may provide a simple means of gaining leads to the functions of many genes for which information is not available currently.

In another perspective, part of the trade-off found among the results obtained from the three protocol analyses could be related to the application, or not, of the background correction before proceeding the normalization methods. Our method of choice – normexp – was clearly better than the traditional subtraction method. Ritchie and others (2007) found normexp as the best background correction method for two-color microarray data, shown to be markedly superior to the usual practice of subtracting local background estimates. Another example of microarray study which applied the normexp background correction in limma is Oshlack et al, (2007) and Martínez-Montemayor et al., (2008).

The false discovery rate (FDR) continues to be a problem that cannot be ignored in our different protocol analyses tested on the same microarray data set. We suggest that accounting for random effects in a mixed model, not establishing a reference for each contrast and applying a more flexible FDR method in small two factorial designs, may generate overcome estimative. Even though, higher than the average found according to the previous analysis where no background correction was done and the *q-value* FDR method was measured, the number of genes identified as differentially expressed were more concise in the both current analyses. The fold change values of the genes were pretty similar between the three protocol analyses, but just few genes with fold change values lower than 1.5 (+/-) were considered as differentially expressed in both current protocols analyses. We suggest that a common FDR controlling method should be tested in these three protocols discussed here to provide further speculations and comparisons, especially with qRT-PCR analyses results already recorded.

The process of comparing and validating different microarray protocol analyses should depend mostly on the size of the experiment, the type of FDR method chosen and in somehow, depending on the experiment design which can leads to clearest interpretations of the results. Straight conclusions about what type of normalization method performs better, is limited in this study, but proceeds background correction based on the normexp method seems to be properly as proved for many others different types of microarray experiments. Anyway, it is always advisable to compare different methods to analyze two-

color microarray experiments in order to check what is best to control the dye bias and allow a clearer identification of differently expressed genes in the experiment.

6. References

Aubert, J, A Bar-Hen, J-J Daudin and S Robin (2004) Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinformatics*, **5**:125 doi:10.1186/1471-2105-5-125.

Beißbarth, T, K Fellenberg, B Brors, R Arribas-Prat, JM Boer, NC Hauser, M Scheideler, JD Hoheisel, G Shutz, A Poustka and M Vingron (2000) Processing and quality control of DNA array Hybridization data. *Bioinformatics*, **16**: 1014-1022.

Bengtsson, A and H Bengtsson (2006) Microarray image analysis: background estimation using quantile and morphological filters. *BMC Bioinformatics*, **6**, 7:96 doi:10.1186/1471-2105-7-96.

Benjamini,Y and Y Hochberg (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Serie B*, **57:** 289-300.

Benjamini,Y and D Yekutieli (2001) The Control of the False Discovery Rate Under Dependency. *Annals of Statistics*, **29:** 1165-1188.

Benjamini,Y, A Krieger and D Yekutieli (2001) Two-Staged Linear Step- Up FDR Controlling Procedure, Department of Statistics and Operation Research, Tel-Aviv University, and Department of Statistics, Wharton School, University of Pennsylvania, Technical Report (Submitted)

Benjamini, Y, E Kenigsberg, A Reiner, D Yekutieli (2005) FDR adjustments of Microarray Experiments. http://www.math.tau.ac.il/~ybenja/Software/fdrame.pdf. Accessed on August 17, 2010

Bilban, M, LK Buehler, S Head, G Desoye and V Quaranta (2002) Normalizing DNA Microarray Data. *Current Issues in Molecular Biology*, **4**: 57-64.

Bolstad, BM, RA Irizarry, MA Strand and TP Speed (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**: 185–193.

Cleveland, WS (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, **74**: 829-836.

Cleveland, WS and SJ Devlin (1998) Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of American Statistical Association*, **83:** 596-610.

Dennis, D Jr, BT Sherman, DA Hosack, J Yang, W Gao, HC Lane and RA Lempicki (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology* **4**, P3.

Degenkolbe, T, PT Do, E Zuther, D Repsilber, D Walther, DK Hincha and KI Kohl (2009) Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. *Plant Molecular Biology*, **69**:133–153.

Draghici S., AK, B Ho and S Shams (2001) Experimental design, analysis of variance and slide quality assessment in gene expression arrays. *Current Opinion in Drug Discovery and Development*, **4** (3):332-337.

Du, P, WA Kibbe, SM Lin (2008) lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, **24:**1547-1548.

Dudoit, S, YH Yang, MJ Callow and TP Speed (2002) Statistical method for identifying with differential expression in replicated cDNA microarray experiments. *Statistica* Sinica, **12**: 111-139.

Eisen, MB, PT Spellman, PO Brown and D Botstein (1998) Cluster analysis and display of genome-wide expression patterns. PNAS Genetics, **95:** 14863–14868.

Fujita, A, JR Sato, LO Rodrigues, CE Ferreira and MC Sogayar (2006) Evaluating different methods of microarray data normalization. *BMC Bioinformatics*, **7**:469 doi:10.1186/1471-2105-7-469.

Gentleman, RC, VJ Carey, DM Bates, B Bolstad, M Dettling, S Dudoit, B Ellis, L Gautier, Y Ge, J Gentry (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, **5**, R80. (based on the statistical programming language R, http://www.R-project.org)

Gilad, Y, Y Gilad, A Oshlack, GK Smyth, TP Speed and KP White (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, **440:** 242–245.

Hartemink, A, D Gifford, TS Jaakkola and RA Young (2001) Maximum likelihood estimation of optimal scaling factors for expression array normalization. In: Microarrays: Optical Technologies and Informatics, edited by Bittner M, Chen Y, Dorsel A, and Doubgherty E. Bellingham, W.A.: SPIE-International Society for Optical Engineering, 2001, p.132–140.

Holm, S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal* of *Statistics*, **6**:65-70.

Huang, da W, BT Sherman and RA Lempicki (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57.

Johnson, RA and DW Wichern (1998) Applied Multivariate Statistical Analisys. 4 th edition. New York: Pretice Hall, 816p.

Kooperberg C, Fazzio T, Delrow J, Tsukiyama T(2002) Improved Background Correction for Spotted DNA Microarrays. *Journal of Computational Biology*, **9:**55-66.

Lehnert, SA, A Reverter, KA Byrne, Y Wang, GS Nattrass, NJ Hudson and PL Greenwood (2007) Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Developmental Biology*, **7**:95. doi:10.1186/1471-213X-7-95.

Liu, L, M Damon, N Guitton, I Guisle, P Ecolan, A Vincent, P Cherel and F Gondret (2009) Differentially-Expressed Genes in Pig Longissimus Muscles with Contrasting Levels of Fat, as Identified by Combined Transcriptomic, Reverse Transcription PCR, and Proteomic Analyses. *Journal* of *Agricultural* and *Food Chemistry*, **57**: 3808–3817.

Lobjois, V, L Liaubet, M SanCristobal, J Glenisson, K Feve, J Rallieres, P Le Roy, D Milan, P Cherel and F Hatey (2008) A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *Animal Genetics* **39**, 147-62.

Martínez-Montemayor, MM, GM Hill, NE Raney, VD Rilington, RJ Tempelman, JE Link, CP Wilkinson, AM Ramos and CW Ernst (2008) Gene expression profiling in hepatic tissue of newly weaned pigs fed pharmacological zinc and phytase supplemented diets. *BMC Genomics* , **9**:421 doi:10.1186/1471-2164-9-421.

McLachlan, GJ, RW Bean, LB Jones (2006) A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics,* **22:**1608-15.

Oshlack, A, AE Chabot, GK Smyth and Y Gilad (2007) Using DNA microarrays to study gene expression in closely related species. *Bioinformatics,* **23**:1235–1242.

Park, T, S-G Yi1, S-H Kang, SY Lee, Y-S Lee and R Simon (2003) Evaluation of normalization methods for microarray data. *BMC Bioinformatics*, **4**:1-13. http://www.biomedcentral.com/1471-2105/4/33.

Peart, MJ, MJ Peart, GK Smyth, RK van Laar, DD Bowtell, VM Richon, PA Marks, AJ Holloway and RW Johnstone (2005) Identification and functional significance of genes regulated by structurally diverse histone deacetylase inhibitors. *PNAS,* **102:** 3697–3702.

Pedotti, P, PAC 't Hoen, E Vreugdenhil, GJ Schenk, RHAM Vossen, Y Ariyurek, M de Hollander, R Kuiper, GJB van Ommen, JT den Dunnen, JM Boer and RX de Menezes (2008) Can subtle changes in gene expression be consistently detected with different microarray platforms? *BMC Genomics,* **9**:124 doi:10.1186/1471-2164-9-124.

Pham, TD, C Wells and DI Crane (2006) Analysis of Microarray Gene Expression Data. *Current Bioinformatics*, ***1:*** *37-53.*

Ponsuksili, S, E Murani, C Walz, M Schwerin and K Wimmers (2007) Pre- and postnatal hepatic gene expression profiles of two pig breeds differing in body composition: insight into pathways of metabolic regulation. *Physiology Genomics* **29**: 267–279.

Reiner, A, D Yekutieli and Y Benjamini (2003) Identifying Differentially Expressed Genes Using False Discovery Rate Controlling Procedures. *Bioinformatics,* **19**: 368-375.

Rainer, J, F Sanchez-Cabo, G Stocker, A Sturn and Z Trajanoski (2006) CARMAweb: comprehensive R- and bioconductor based web service for microarray data analysis. *Nucleic Acids Research,* **34.** doi:10.1093/nar/gkl038.

Ritchie, ME, J Silver, A Oshlack, M Holmes, D Diyagama, A Holloway, GK Smyth (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics,* 23:2700–2707.

SAS Systems: **The SAS System for Windows, version 9.2.** SAS Institute, Inc., Cary, NC; 2008.

Silver, JD, ME Ritchie, GK Smyth (2008) Microarray background correction: maximum likelihood estimation for the normal–exponential convolution. *Biostatistics,* pp. 1–12. doi:10.1093/biostatistics/kxn042.

Shi, W, A Banerjee, ME Ritchie, S Gerondakis and GK Smyth (2009) Illumina WG-6 BeadChip strips should be normalized separately. *BMC Bioinformatics* 2009, **10**:372 doi:10.1186/1471-2105-10-372.

Smith, M, R Kohn (1996) Nonparametric regression using Bayesian variable Selection. *Journal of Econometrics*, **75:** 317 343.

Smyth, GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology, **3,** Article 3.*

Smyth GK (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* Edited by: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W. New York: Springer, 397-420.

Soler, JMP, FHFP da Rosa, S Chiavegatto, I Aneas and JE Krieger (2004) Use of splines for normalization of microarray gene expression data. *Bioscience Journal* Especial, p. 101-116.

Steibel, JP, M Wysocki, LK Lunney, AM Ramos, Z-T Hu, MF Rothschild and CW Ernst (2009) Assessment of the swine protein-annotated oligonucleotide microarray. *Animal Genetics* **40**, 883-93.

Storey J.D. & Tibshirani R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440-45.

Sharma, S (1996) Applied multivariate techniques, John Wiley & Sons, Inc. 229p.

Tempelman, RJ (2008) Statistical Analysis of Efficient Unbalanced Factorial Designs for Two-ColorMicroarray Experiments. *International Journal of Plant Genomics*, Article ID 584360, 16 pages. doi:10.1155/2008/584360.

Verducci, JS, VF Melfi, S Lin, Z Wang, S Roy and C K Sen (2006) Microarray analysis of gene expression: considerations in data mining and statistical treatment. *Physiol Genomics* **25**: 355–363.

Vinciotti, V, R Khanin, D D'Alimonte, X Liu, N Cattini, G Hotchkiss, G Bucca, O de Jesus, J Rasaiyaah, CP Smith, P Kellam, E Wit (2005) An experimental evaluation of a loop versus a reference design for two-channel microarrays. *Bioinformatics,* **21**(4):492-501.

Wegman EJ and IW Wright (1983) Splines in Statistics. *Journal of the American Statistical Association*, **78**: 351- 365.

Weniger, M, JC Engelmann and J Schultz (2007) Genome Expression Pathway Analysis Tool – Analysis and visualization of microarray gene expression data under genomic, proteomic and metabolic context. *BMC Bioinformatics*, **8**:179 doi:10.1186/1471-2105-8-179.

Workman, C, LJ Jensen, H Jarmer, R Berka, L Gautier, HB Nielsen, H-H Saxild, C Nielsen, S Brunak and S Knudsen (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology,* **3**(9):research0048.1–0048.16.

Xu, W and M Faisal (2010) Factorial microarray analysis of zebra mussel (Dreissena polymorpha: Dreissenidae, Bivalvia) adhesion. *BMC Genomics*, **11**:341 http://www.biomedcentral.com/1471-2164/11/341.

Yang, YH, S Dudoit, P Luu, DM Lin, V Peng, J Ngai and TP Speed (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, **30:** 1-15.

Yang, YH and NP Thorne (2003) Normalization for Two-Color cDNA Microarray Data. *Lecture Notes-Monograph Series*, **40**, Statistics and Science: A Festschrift for Terry Speed (2003), pp. 403-418 (Published by: Institute of Mathematical StatisticsStable URL: http://www.jstor.org/stable/4356198)

Yekutieli D and Y Benjamini (1999) Resampling-Based False Discovery Rate Controlling Multiple Test Procedures for Correlated Test Statistics. *Journal of Statistical Planning and Inference*, **82**: 171-196.

Zhao, S-H., D Nettleton, W Liu, C Fitzsimmons, CW Ernst, NE Raney and CK Tuggle (2003) Complementary DNA macroarray analyses of differential gene expression in porcine fetal and postnatal muscle. *Journal of Animal Science* **81:** 2179-88.

Zhao Y, M-C Li and R Simon (2005) An adaptive method for cDNA microarray normalization. *BMC Bioinformatics*, **6**:28 doi:10.1186/1471-2105-6-28.

General conclusions

On the whole this research provides new data regarding developmental changes of expression profile in different functional pathways related to myogenic process and also to an early postnatal period of the muscle development of pigs. Started with a transcriptional global overview provided by a two-color microarray experiment, twenty seven genes specifically investigated by qRT-PCR analyses, increased the knowledge of molecular events involved in skeletal muscle development of pigs. Both qRT-PCR methodologies applied, successfully described the molecular events accompanying skeletal muscle development in pigs and a new reference gene for expression profile investigation of pigs is proposed DDIT3 (DNA-damage-inducible transcript 3).

We have identified both developmental and breed-specific patterns of gene expression comparing the Brazilian local pig breed Piau with a crossbred from United States and with a composite line developed in Brazil. Our first evaluation, based on microarray analysis of genetic groups at 40 and 70 d of gestation, suggested that fetuses from a more highly muscled breed of pigs show a later peak of myogenesis-related gene expression during fetal development. At that time, most of the genes specifically analyzed by qRT-PCR were related to muscle fiber structure and formation. However, the second analysis investigated others fourteen genes expressions through the five time-points of the development more related to signaling pathways acting on cell proliferation, differentiation and energy metabolism. At this time, mechanisms of activation and maintenance of myoblast differentiation stages during the muscle development of white composite genetic pigs selected for lean muscle mass were more active. Also, intrinsic mechanisms of myoblast proliferation differentially acting between these two groups of pigs still need to be understood. Anyway, two interesting mechanisms that may interfere in the differences of muscle development between both genetic groups are proposed in this research: the ubiquitin proteasome system (proteolysis pathway) and the Wnt/β-catenin signaling pathway.

The necessity of comparisons of different protocol analyses of two-color microarray experiments is also advisable in order to check what is best to control the dye bias and to allow a clearer identification of differently expressed genes in the experiment. The Robust Spline normalization method, which is

recognized by its flexibility, can results in more accurate normalized data together with the application of the normexp background correction method, generating consistent results.

This study is the first report to evaluate a transcriptome overview of the Brazilian native Piau pig breed using the microarray methodology, and added intrinsic information about determinant mechanisms for muscle mass content of pigs. In this way, future hypothesis-based testing and novel strategies are proposed to enhance the efficiency of lean tissue deposition in the genetic improvement of this specie.

Appendix 1

Table S1 Genes selected for qRT-PCR.

Gene Name	Symbol	TaqMan® Assay/Accession No[1].
carbonic anhydrase III, muscle specific	CA3	Ss03391841_m1
catenin (cadherin-associated protein), beta 1	CTNNB1	Ss03394813_g1
cathepsin L2	CTSL2	Ss03392489_u1
delta-like 1 homolog	DLK1	Ss03380929_u1
F-box protein 32	FBXO32	Ss03388469_m1
hypoxanthine phosphoribosyltransferase 1	HPRT1	Ss03388273_m1
myozenin 1	MYOZ1	Ss03391954_m1
nebulin-related anchoring protein	NRAP	Ss03388532_m1
ornithine decarboxylase 1	ODC1	NM_001122983.1
sarcopilin	SLN	Ss03380464_u1
signal transducer and activator of transcription 1	STAT1	Ss03392290_m1
TIMP metallopeptidase inhibitor 3	TIMP3	AF156031
tenascin C	TNC	Ss03394416_m1
ubiquitin specific peptidase 13	USP13	BX668182

[1]Applied Biosystems TaqMan® assay product number or GenBank accession number of sequence submitted for Custom TaqMan® Gene Expression Assay design.

Appendix 2

Supplementary table 1. Gene Symbol (HGNC), primer sequence, Reference sequence (NCBI), species of reference and efficiency of amplification of each primer of target and reference gene.

Gene Symbol	Primer sequence	NCBI Ref. Sequence	Specie	Efficiency
CAV3	F- AGTGAAGGTGGACTTCGAGGATGT	NM_001037149.1	*Sus Scrofa*	0.98
	R- ATACTTGGACACGGTGAAGGTGGT			
CSDA	F- TTACCACGTGGGACAGACCTTTGA	NM_003651.4	*Sus Scrofa*	0.99
	R- TTCGATGAACCGGTCCCTGAAGTT			
CTNNB1	F- TATCCAGTTGATGGGCTGCCAGAT	NM_214367.1	*Sus Scrofa*	1.26
	R- ACAGGTCAGTATCAAACCAGGCCA			
DLK1	F- TCAGGCCATCTGCTTTACCATCCT	NM_001048187.1	*Sus Scrofa*	0.88
	R- GGTTCTTCTTCTTGCGCAGCATGT			
DDIT3	F- GAACGAACGGCTCAAGCAGGAAAT	NM_001144845.1	*Sus Scrofa*	1.07
	R- TTGGTAGCCACTTCCAGGAAAGGT			
FABP4	F- AGATTGCCTTCAAATTGGGCCAGG	NM_019102.2	*Sus Scrofa*	0.8
	R- TCTTTCCATCCCACTTCTGCACCT			
GAPDH	F- CAAAGTGGACATTGTCGCCATCA	AF017079.1	*Sus Scrofa*	1.01
	R- AGCTTCCCATTCTCAGCCTTGACT			
GRN	F- CCCATGGCAAAGATCCCTTCACAA	NM_001044578.1	*Sus Scrofa*	0.84
	R- TGCAGCAGGTGGAATTATTGTGGC			
HOXA5	F- AGATCATAGTTCCGTGAGCGAGCA	NM_019102.2	*H. sapiens*	0.88
	R- TGAGATCCATGCCATTGTAGCCGT			
HSP90B1	F- ATGTCACAGACACTGGTGTGGGAA	NM_214103.1	*Sus Scrofa*	0.85
	R- AGGCAGAATAGAAGCCAACACCGA			
MSTN	F- ACTTCGCCTGGAAACAGCTCCTAA	NM_214435.2	*Sus Scrofa*	0.8
	R- ATGATCGTTTCCGTCGTAGCGTGA			
PPARGC1A	F- AAACCCACAGAGACCCGAAACAGT	NM_213963.1	*Sus Scrofa*	1.32
	R- GGCTTGTAAATGTTGCGACTGCGA			
PAK1	F- GGCCAAGCTCTCTGCAAATGGAAA	NM_002576.4	*H. sapiens*	1.03
	R- GCTGACATGACAAGCCACAATGCT			
TNC	F- ACTGCTCCCAAGCAATGCTGAATG	NM_214230	*Sus Scrofa*	0.85
	R- AGAAGACTTCCAGCTTCTGGGCTT			
TRIM63	F- TGCTGGTGGAGAACATCATCGACA	NM_032588.2	*H. sapiens*	0.85
	R- TCTCATCTTCGTGCTCCTTGCACA			

Supplementary table 2. Statistical analysis results for each contrast including Genetic group comparison (breed) and Age comparisons (age) for each target gene.

Contrast	Estimate	Fold change	StdErr	P value
breed:c-p|age=21d|gene:CAV3	0.8275	-1.7746	0.7846	0.304139
breed:c-p|age=40d|gene:CAV3	-0.0570	1.0403	0.7846	0.942806
breed:c-p|age=70d|gene:CAV3	-1.0425	2.0597	0.7846	0.198900
breed:c-p|age=90d|gene:CAV3	-1.4048	2.6478	0.7846	0.088517
breed:c-p|age=10wk|gene:CAV3	-0.9058	1.8736	0.7846	0.261889
age:21d-40d|breed=c|gene:CAV3	6.8840	-118.1110	0.7846	0.000000027
age:40d-70d|breed=c|gene:CAV3	1.5247	-2.8772	0.7846	0.066191
age:70d-90d|breed=c|gene:CAV3	0.4028	-1.3221	0.7846	0.613273
age:90d-10wk|breed=c|gene:CAV3	-2.5243	5.7531	0.7846	0.004321
age:21d-40d|breed=p|gene:CAV3	5.9995	-63.9778	0.7846	2.3254E-07
age:40d-70d|breed=p|gene:CAV3	0.5392	-1.4531	0.7846	0.499849
age:70d-90d|breed=p|gene:CAV3	0.0405	-1.0285	0.7846	0.959344
age:90d-10wk|breed=p|gene:CAV3	-2.0253	4.0785	0.7846	0.017829
breed:c-p|age=21d|gene:CSDA	-0.2325	1.1749	0.6560	0.726732
breed:c-p|age=40d|gene:CSDA	-0.4685	1.3837	0.6560	0.483369
breed:c-p|age=70d|gene:CSDA	-1.1323	2.1921	0.6560	0.099742
breed:c-p|age=90d|gene:CSDA	-3.1470	8.8581	0.6560	0.000109697
breed:c-p|age=10wk|gene:CSDA	-0.7053	1.6305	0.6560	0.295083
age:21d-40d|breed=c|gene:CSDA	1.0185	-2.0258	0.6560	0.136201
age:40d-70d|breed=c|gene:CSDA	1.8648	-3.6423	0.6560	0.010058
age:70d-90d|breed=c|gene:CSDA	3.6760	-12.7816	0.6560	1.74479E-05
age:90d-10wk|breed=c|gene:CSDA	-3.7338	13.3044	0.6560	1.43265E-05
age:21d-40d|breed=p|gene:CSDA	0.7825	-1.7201	0.6560	0.246892
age:40d-70d|breed=p|gene:CSDA	1.2010	-2.2990	0.6560	0.082073
age:70d-90d|breed=p|gene:CSDA	1.6613	-3.1631	0.6560	0.019806
age:90d-10wk|breed=p|gene:CSDA	-1.2922	2.4489	0.6560	0.062871
breed:c-p|age=21d|gene:CTNNB1	0.3016	-1.2325	0.7582	0.695303
breed:c-p|age=40d|gene:CTNNB1	-1.0892	2.1275	0.6824	0.126964
breed:c-p|age=70d|gene:CTNNB1	-2.6857	6.4338	0.6824	0.000886
breed:c-p|age=90d|gene:CTNNB1	-1.2450	2.3702	0.6824	0.083843
breed:c-p|age=10wk|gene:CTNNB1	-0.2198	1.1646	0.6824	0.750869
age:21d-40d|breed=c|gene:CTNNB1	-0.3755	1.2973	0.6824	0.588564
age:40d-70d|breed=c|gene:CTNNB1	-1.3683	2.5817	0.6824	0.059391
age:70d-90d|breed=c|gene:CTNNB1	-3.4665	11.0540	0.6824	0.000066
age:90d-10wk|breed=c|gene:CTNNB1	0.2738	-1.2090	0.6824	0.692703
age:21d-40d|breed=p|gene:CTNNB1	-1.7663	3.4018	0.7582	0.031298
age:40d-70d|breed=p|gene:CTNNB1	-2.9648	7.8074	0.6824	0.000348
age:70d-90d|breed=p|gene:CTNNB1	-2.0258	4.0723	0.6824	0.007882
age:90d-10wk|breed=p|gene:CTNNB1	1.2990	-2.4606	0.6824	0.072216

Supplementary table 2 cont. Statistical analysis results for each contrast including Genetic group comparison (breed) and Age comparisons (age) for each target gene.

Contrast	Estimate	Fold change	StdErr	P value
breed:c-p|age=21d|gene:DLK1	0.9558	-1.9397	1.2991	0.470472
breed:c-p|age=40d|gene:DLK1	-0.5230	1.4369	1.2991	0.691554
breed:c-p|age=70d|gene:DLK1	-9.0874	543.9650	1.2999	8.89119E-07
breed:c-p|age=90d|gene:DLK1	-2.0872	4.2491	1.2991	0.123899
breed:c-p|age=10wk|gene:DLK1	-3.4115	10.6405	1.2991	0.016232
age:21d-40d|breed=c|gene:DLK1	5.2628	-38.3946	1.2991	0.000630
age:40d-70d|breed=c|gene:DLK1	0.9935	-1.99101	1.2991	0.453403
age:70d-90d|breed=c|gene:DLK1	-0.6065	1.52256	1.2991	0.645681
age:90d-10wk|breed=c|gene:DLK1	-8.7555	432.18345	1.2991	1.51767E-06
age:21d-40d|breed=p|gene:DLK1	3.7840	-13.77519	1.2991	0.008636
age:40d-70d|breed=p|gene:DLK1	-7.5709	190.13364	1.2999	0.000011
age:70d-90d|breed=p|gene:DLK1	6.3937	-84.08098	1.2999	0.000083
age:90d-10wk|breed=p|gene:DLK1	-10.0798	1082.26144	1.2991	1.92656E-07
breed:c-p|age=21d|gene:FABP4	-0.2718	1.2073	0.8318	0.747208
breed:c-p|age=40d|gene:FABP4	0.4622	-1.3776	0.8318	0.584623
breed:c-p|age=70d|gene:FABP4	-1.9835	3.9545	0.8318	0.027110
breed:c-p|age=90d|gene:FABP4	-1.1632	2.2395	0.8318	0.177306
breed:c-p|age=10wk|gene:FABP4	-0.0513	1.0362	0.8318	0.951402
age:21d-40d|breed=c|gene:FABP4	3.0495	-8.2792	0.8318	0.001533
age:40d-70d|breed=c|gene:FABP4	3.1750	-9.0317	0.8318	0.001079
age:70d-90d|breed=c|gene:FABP4	1.2497	-2.3779	0.8318	0.148619
age:90d-10wk|breed=c|gene:FABP4	-0.5752	1.4898	0.8318	0.497207
age:21d-40d|breed=p|gene:FABP4	3.7835	-13.7704	0.8318	0.000195
age:40d-70d|breed=p|gene:FABP4	0.7293	-1.6579	0.8318	0.390987
age:70d-90d|breed=p|gene:FABP4	2.0700	-4.1989	0.8318	0.021755
age:90d-10wk|breed=p|gene:FABP4	0.5367	-1.4506	0.8318	0.526129
breed:c-p|age=21d|gene:GAPDH	-0.0512	1.0361	1.1351	0.964494
breed:c-p|age=40d|gene:GAPDH	-0.3798	1.3012	1.1351	0.741395
breed:c-p|age=70d|gene:GAPDH	-2.1339	4.3892	1.1605	0.079871
breed:c-p|age=90d|gene:GAPDH	-3.2480	9.5005	1.1351	0.009660
breed:c-p|age=10wk|gene:GAPDH	-1.5443	2.9167	1.1351	0.188824
age:21d-40d|breed=c|gene:GAPDH	-0.6228	1.5399	1.1351	0.589290
age:40d-70d|breed=c|gene:GAPDH	0.8448	-1.7960	1.1605	0.474552
age:70d-90d|breed=c|gene:GAPDH	3.1712	-9.0081	1.1605	0.012344
age:90d-10wk|breed=c|gene:GAPDH	0.2625	-1.1996	1.1351	0.819467
age:21d-40d|breed=p|gene:GAPDH	-0.9515	1.9339	1.1351	0.411809
age:40d-70d|breed=p|gene:GAPDH	-0.9093	1.8782	1.1351	0.432499
age:70d-90d|breed=p|gene:GAPDH	2.0572	-4.1617	1.1351	0.085003
age:90d-10wk|breed=p|gene:GAPDH	1.4412	-2.7154	1.1351	0.218811

Supplementary table 2 cont. Statistical analysis results for each contrast including Genetic group comparison (breed) and Age comparisons (age) for each target gene.

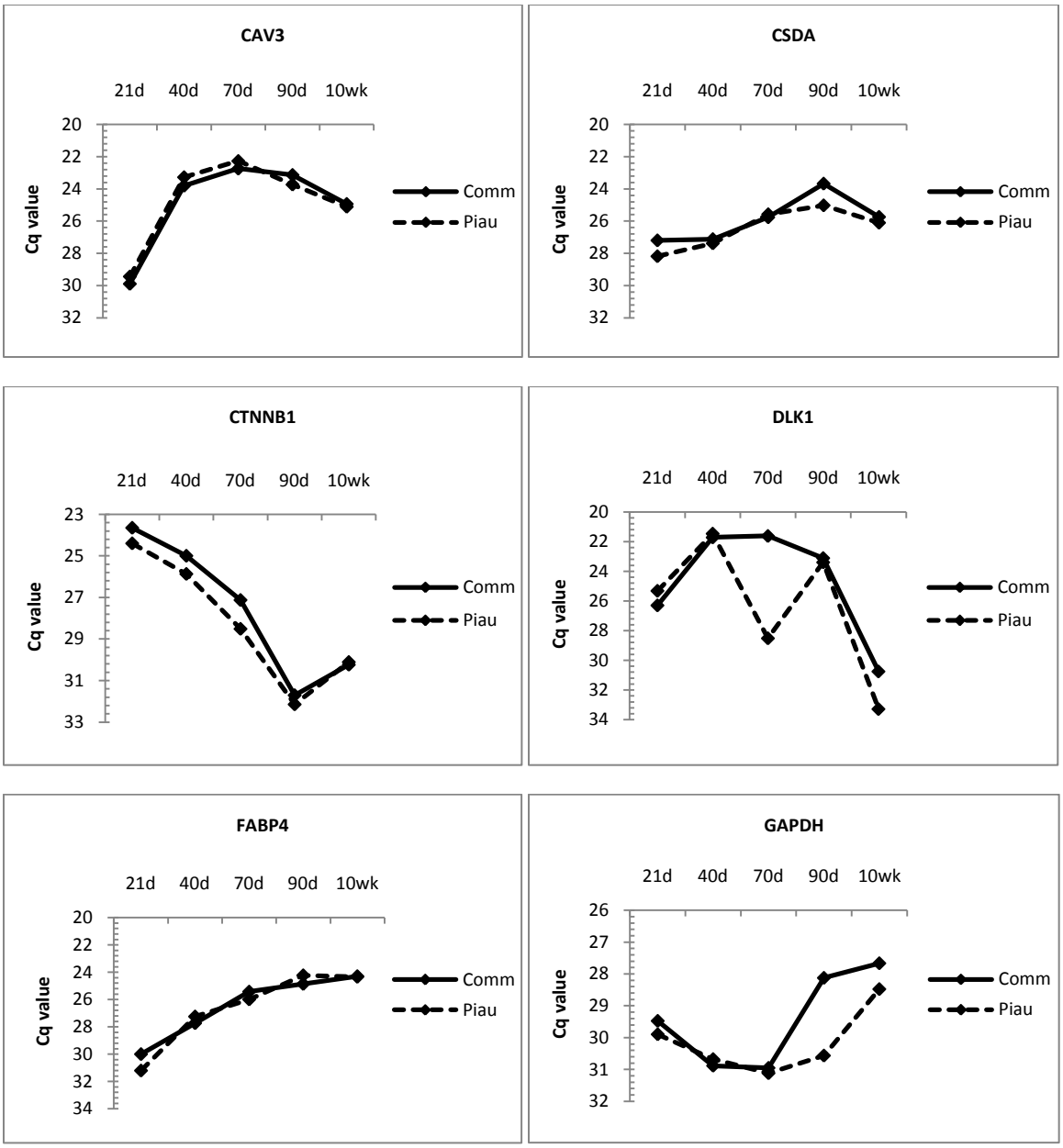
Contrast	Estimate	Fold change	StdErr	P value
breed:c-p|age=21d|gene:GRN	-1.1813	2.2679	1.2023	0.337548
breed:c-p|age=40d|gene:GRN	-0.9887	1.9844	1.2023	0.420582
breed:c-p|age=70d|gene:GRN	-1.5563	2.9410	1.2023	0.210240
breed:c-p|age=90d|gene:GRN	-1.2772	2.4236	1.2023	0.300759
breed:c-p|age=10wk|gene:GRN	-0.4887	1.4031	1.2023	0.688725
age:21d-40d|breed=c|gene:GRN	-1.1392	2.2025	1.2023	0.354685
age:40d-70d|breed=c|gene:GRN	-1.3695	2.5838	1.2023	0.268122
age:70d-90d|breed=c|gene:GRN	1.8833	-3.6893	1.2023	0.132921
age:90d-10wk|breed=c|gene:GRN	-4.8177	28.2008	1.2023	0.000692
age:21d-40d|breed=p|gene:GRN	-0.9465	1.9272	1.2023	0.440358
age:40d-70d|breed=p|gene:GRN	-1.9372	3.8295	1.2023	0.122792
age:70d-90d|breed=p|gene:GRN	2.1625	-4.4769	1.2023	0.087179
age:90d-10wk|breed=p|gene:GRN	-4.0292	16.3268	1.2023	0.003179
breed:c-p|age=21d|gene:HOXA5	2.2935	-4.9024	1.2557	0.083543
breed:c-p|age=40d|gene:HOXA5	-2.2272	4.6821	1.2557	0.092159
breed:c-p|age=70d|gene:HOXA5	-1.7432	3.3476	1.2557	0.181145
breed:c-p|age=90d|gene:HOXA5	-0.7235	1.6511	1.2557	0.571261
breed:c-p|age=10wk|gene:HOXA5	-1.8614	3.6336	1.4029	0.200278
age:21d-40d|breed=c|gene:HOXA5	1.1987	-2.2953	1.2557	0.351780
age:40d-70d|breed=c|gene:HOXA5	-1.1525	2.2230	1.2557	0.370222
age:70d-90d|breed=c|gene:HOXA5	0.3893	-1.3098	1.2557	0.759895
age:90d-10wk|breed=c|gene:HOXA5	-3.7186	13.1645	1.4029	0.015767
age:21d-40d|breed=p|gene:HOXA5	-3.3220	10.0005	1.2557	0.015958
age:40d-70d|breed=p|gene:HOXA5	-0.6685	1.5894	1.2557	0.600639
age:70d-90d|breed=p|gene:HOXA5	1.4090	-2.6555	1.2557	0.275812
age:90d-10wk|breed=p|gene:HOXA5	-4.8565	28.9702	1.2557	0.001038
breed:c-p|age=21d|gene:HSP90B1	0.9345	-1.9112	0.6627	0.173831
breed:c-p|age=40d|gene:HSP90B1	-0.2300	1.1728	0.6627	0.732149
breed:c-p|age=70d|gene:HSP90B1	-0.1847	1.1365	0.6627	0.783352
breed:c-p|age=90d|gene:HSP90B1	0.0267	-1.0186	0.6627	0.968299
breed:c-p|age=10wk|gene:HSP90B1	0.1822	-1.1346	0.6627	0.786208
age:21d-40d|breed=c|gene:HSP90B1	-0.0140	1.0098	0.6627	0.983353
age:40d-70d|breed=c|gene:HSP90B1	-1.3348	2.5225	0.6627	0.057615
age:70d-90d|breed=c|gene:HSP90B1	0.3532	-1.2774	0.6627	0.599934
age:90d-10wk|breed=c|gene:HSP90B1	-3.3783	10.3987	0.6627	0.000055
age:21d-40d|breed=p|gene:HSP90B1	-1.1785	2.2634	0.6627	0.090534
age:40d-70d|breed=p|gene:HSP90B1	-1.2895	2.4444	0.6627	0.065846
age:70d-90d|breed=p|gene:HSP90B1	0.5645	-1.4789	0.6627	0.404365
age:90d-10wk|breed=p|gene:HSP90B1	-3.2228	9.3362	0.6627	0.000094

Supplementary table 2 cont. Statistical analysis results for each contrast including Genetic group comparison (breed) and Age comparisons (age) for each target gene.

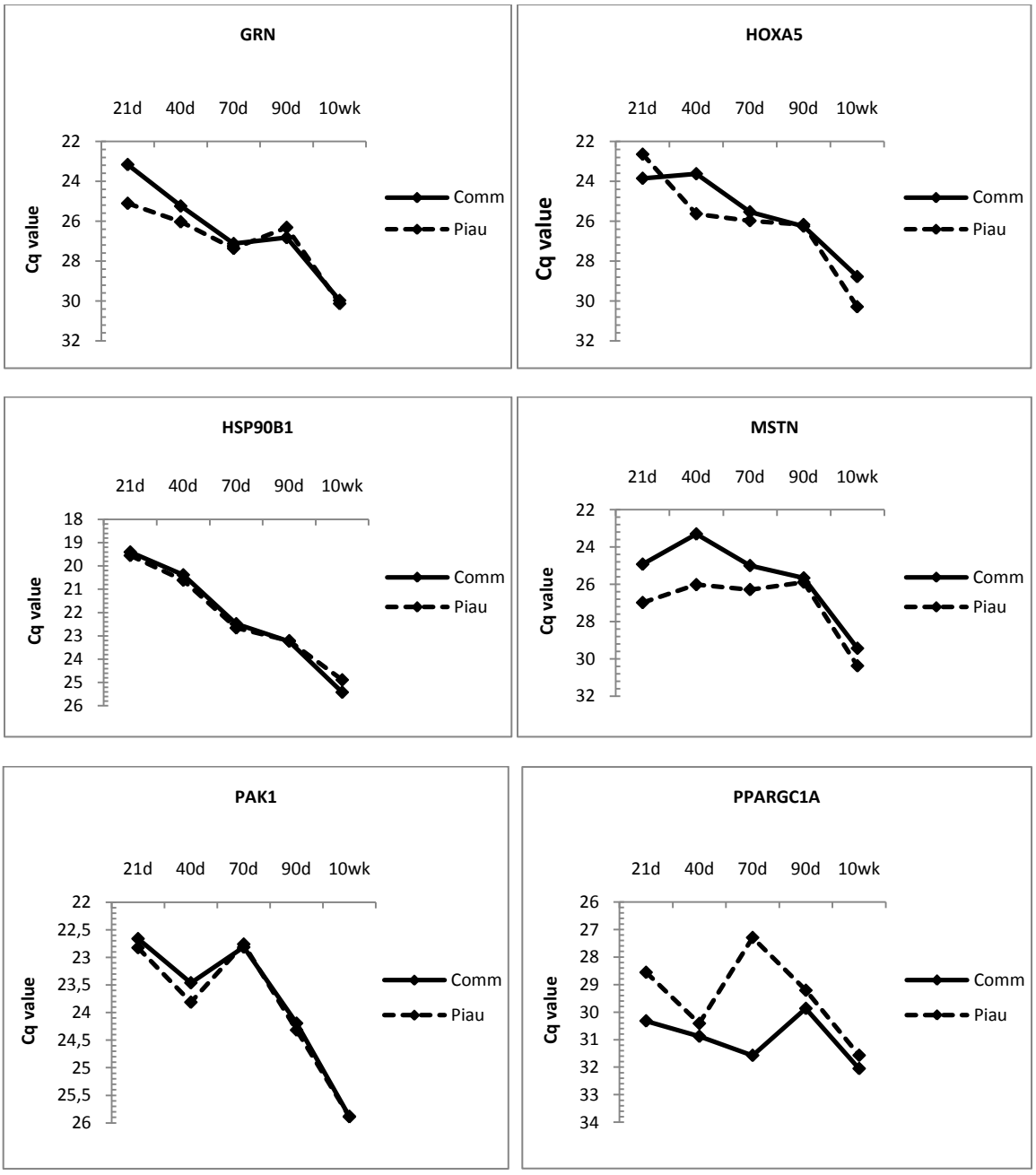
Contrast	Estimate	Fold change	StdErr	P value
breed:c-p|age=21d|gene:MSTN	-2.0620	4.1756	0.8821	0.029922
breed:c-p|age=40d|gene:MSTN	-2.7423	6.6915	0.8821	0.005534
breed:c-p|age=70d|gene:MSTN	-2.6985	6.4913	0.8821	0.006194
breed:c-p|age=90d|gene:MSTN	-1.0245	2.0343	0.8821	0.112686
breed:c-p|age=10kw|gene:MSTN	-1.3080	2.4760	0.8821	0.153727
age:21d-40d|breed=c|gene:MSTN	1.4668	-2.7641	0.8821	0.111944
age:40d-70d|breed=c|gene:MSTN	-0.8293	1.7769	0.8821	0.358366
age:70d-90d|breed=c|gene:MSTN	0.0282	-1.0197	0.8821	0.974845
age:90d-10wk|breed=c|gene:MSTN	-4.5535	23.4823	0.8821	0.000047
age:21d-40d|breed=p|gene:MSTN	0.7865	-1.7249	0.8821	0.383220
age:40d-70d|breed=p|gene:MSTN	-0.7855	1.7237	0.8821	0.383814
age:70d-90d|breed=p|gene:MSTN	0.7022	-1.6269	0.8821	0.435391
age:90d-10wk|breed=p|gene:MSTN	-3.8370	14.2907	0.8821	0.000311
breed:c-p|age=21d|gene:PAK1	0.1805	-1.1333	0.5571	0.749289
breed:c-p|age=40d|gene:PAK1	-1.1210	2.1750	0.5571	0.057850
breed:c-p|age=70d|gene:PAK1	0.0525	-1.0371	0.5571	0.925854
breed:c-p|age=90d|gene:PAK1	-0.1242	1.0899	0.5571	0.825879
breed:c-p|age=10wk|gene:PAK1	-0.8810	1.8417	0.5571	0.129451
age:21d-40d|breed=c|gene:PAK1	-0.1130	1.0815	0.5571	0.841304
age:40d-70d|breed=c|gene:PAK1	1.5413	-2.9106	0.5571	0.011896
age:70d-90d|breed=c|gene:PAK1	-0.4957	1.4100	0.5571	0.384164
age:90d-10wk|breed=c|gene:PAK1	-2.8037	6.9821	0.5571	0.000064
age:21d-40d|breed=p|gene:PAK1	-1.0535	2.0756	0.5571	0.073179
age:40d-70d|breed=p|gene:PAK1	2.7148	-6.5652	0.5571	0.000092
age:70d-90d|breed=p|gene:PAK1	-0.6723	1.5936	0.5571	0.241547
age:90d-10wk|breed=p|gene:PAK1	-3.5605	11.7982	0.5571	0.000003
breed:c-p|age=21d|gene:PPARGC1A	2.1392	-4.4051	1.2269	0.096588
breed:c-p|age=40d|gene:PPARGC1A	-0.1130	1.0815	1.2269	0.927533
breed:c-p|age=70d|gene:PPARGC1A	2.7695	-6.8187	1.2269	0.035327
breed:c-p|age=90d|gene:PPARGC1A	-0.1417	1.1032	1.2269	0.909226
breed:c-p|age=10wk|gene:PPARGC1A	-0.2458	1.1858	1.2269	0.843214
age:21d-40d|breed=c|gene:PPARGC1A	0.2262	-1.1697	1.2269	0.855602
age:40d-70d|breed=c|gene:PPARGC1A	-0.2317	1.1742	1.2269	0.852133
age:70d-90d|breed=c|gene:PPARGC1A	2.4983	-5.6503	1.2269	0.055179
age:90d-10wk|breed=c|gene:PPARGC1A	-2.9022	7.4755	1.2269	0.028223
age:21d-40d|breed=p|gene:PPARGC1A	-2.0260	4.0727	1.2269	0.114287
age:40d-70d|breed=p|gene:PPARGC1A	2.6508	-6.2803	1.2269	0.043036
age:70d-90d|breed=p|gene:PPARGC1A	-0.4128	1.3313	1.2269	0.740008
age:90d-10wk|breed=p|gene:PPARGC1A	-3.0063	8.0352	1.2269	0.023599

Supplementary table 2 cont. Statistical analysis results for each contrast including Genetic group comparison (breed) and Age comparisons (age) for each target gene.

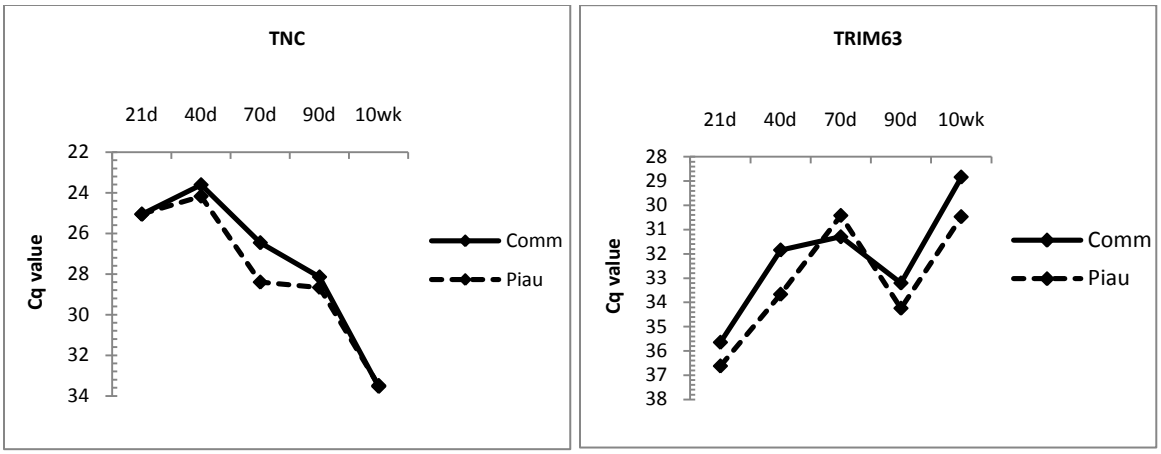
Contrast	Estimate	Fold change	StdErr	P value
breed:c-p|age=21d|gene:TNC	0.0038	-1.0027	0.8211	0.996321
breed:c-p|age=40d|gene:TNC	-1.3293	2.5129	0.8211	0.121128
breed:c-p|age=70d|gene:TNC	-4.1225	17.4179	0.8211	0.000066
breed:c-p|age=90d|gene:TNC	-2.3060	4.9451	0.8211	0.010855
breed:c-p|age=10wk|gene:TNC	-0.8262	1.7730	0.8211	0.326367
age:21d-40d|breed=c|gene:TNC	2.1413	-4.4117	0.8211	0.016840
age:40d-70d|breed=c|gene:TNC	-1.9590	3.8879	0.8211	0.027047
age:70d-90d|breed=c|gene:TNC	-0.8027	1.7443	0.8211	0.339988
age:90d-10wk|breed=c|gene:TNC	-6.5108	91.1919	0.8211	1.337E-07
age:21d-40d|breed=p|gene:TNC	0.8082	-1.7510	0.8211	0.336766
age:40d-70d|breed=p|gene:TNC	-4.7522	26.9491	0.8211	0.000012
age:70d-90d|breed=p|gene:TNC	1.0138	-2.0193	0.8211	0.231257
age:90d-10wk|breed=p|gene:TNC	-5.0310	32.6950	0.8211	0.000005
breed:c-p|age=21d|gene:TRIM63	-0.2096	1.1564	0.9616	0.829583
breed:c-p|age=40d|gene:TRIM63	-2.0233	4.0652	0.8774	0.032334
breed:c-p|age=70d|gene:TRIM63	-0.4440	1.3604	0.8774	0.618550
breed:c-p|age=90d|gene:TRIM63	-2.8347	7.1338	0.8774	0.004326
breed:c-p|age=10wk|gene:TRIM63	-1.9758	3.9335	0.8774	0.036130
age:21d-40d|breed=c|gene:TRIM63	4.7465	-26.8440	0.9616	0.000076
age:40d-70d|breed=c|gene:TRIM63	1.0538	-2.0760	0.8774	0.244223
age:70d-90d|breed=c|gene:TRIM63	-0.3152	1.2442	0.8774	0.723335
age:90d-10wk|breed=c|gene:TRIM63	2.6937	-6.4696	0.8774	0.006209
age:21d-40d|breed=p|gene:TRIM63	2.9328	-7.6361	0.8774	0.003357
age:40d-70d|breed=p|gene:TRIM63	2.6332	-6.2039	0.8774	0.007243
age:70d-90d|breed=p|gene:TRIM63	-2.7058	6.5243	0.8774	0.006019
age:90d-10wk|breed=p|gene:TRIM63	3.5525	-11.7330	0.8774	0.000664



Supplementary figure 1- Quantitative cycle of each genetic group in each pre and postnatal age for each target gene.



Supplementary figure 1 cont.- Quantitative cycle of each genetic group in each pre and postnatal age for each target gene.



Supplementary figure 1 cont.- Quantitative cycle of each genetic group in each pre and postnatal age for each target gene.

Appendix 3

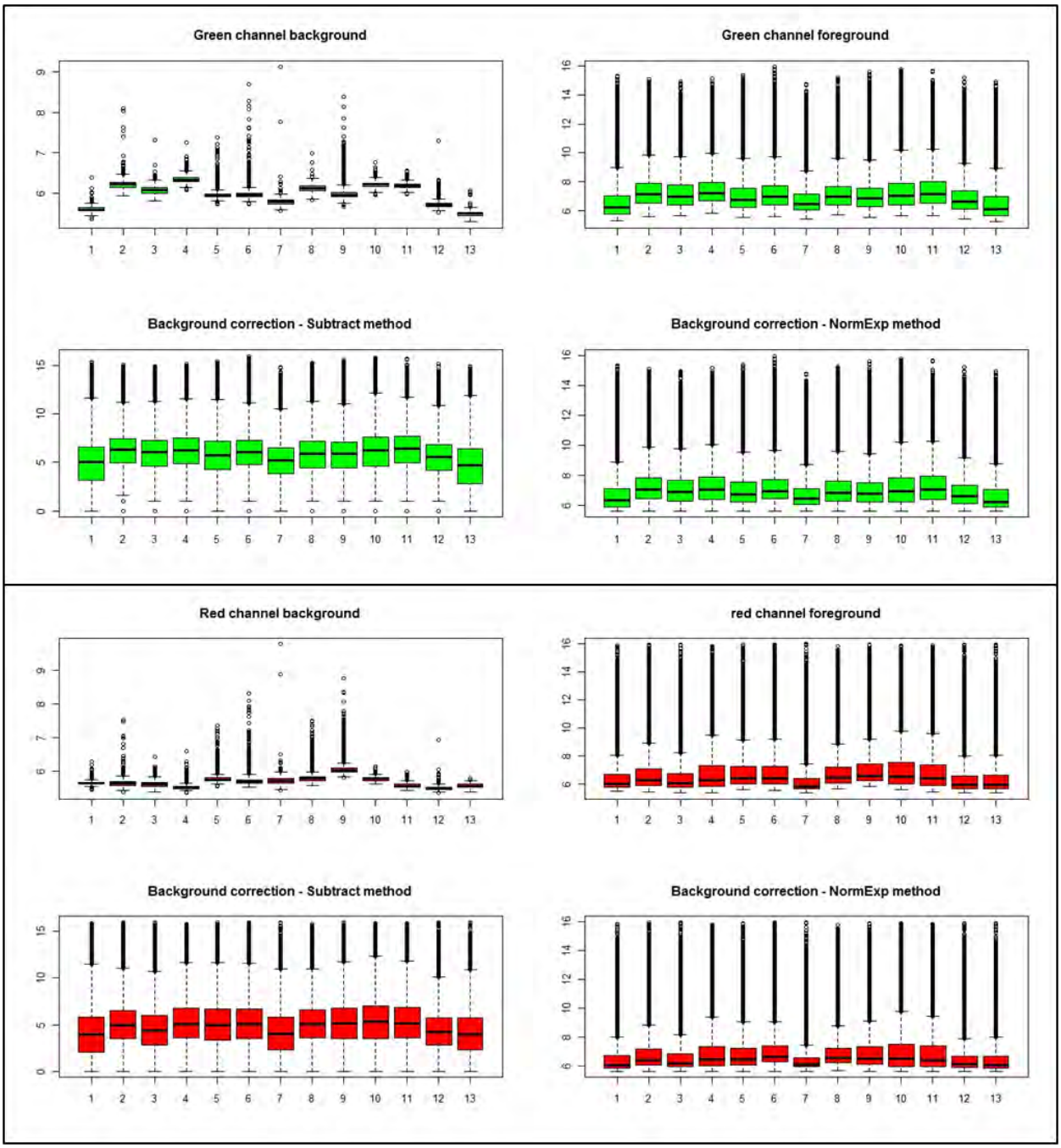


Figure1- Boxplot representing the background, foreground, background correction according to the Subtract method and background correction according to the Normexp method for green and red channel fluorescence in each of the 13 slides.

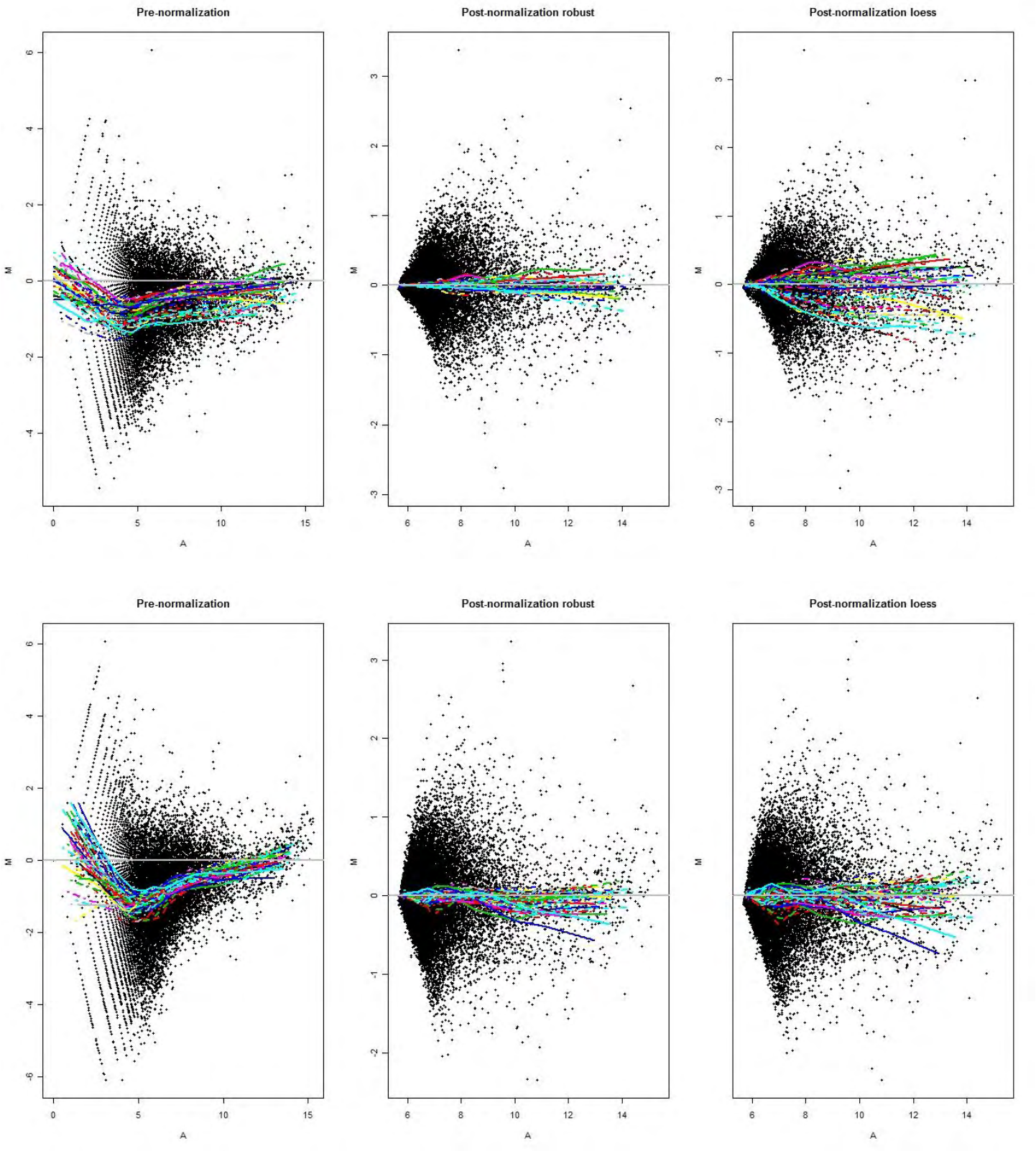


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 1 and 2, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

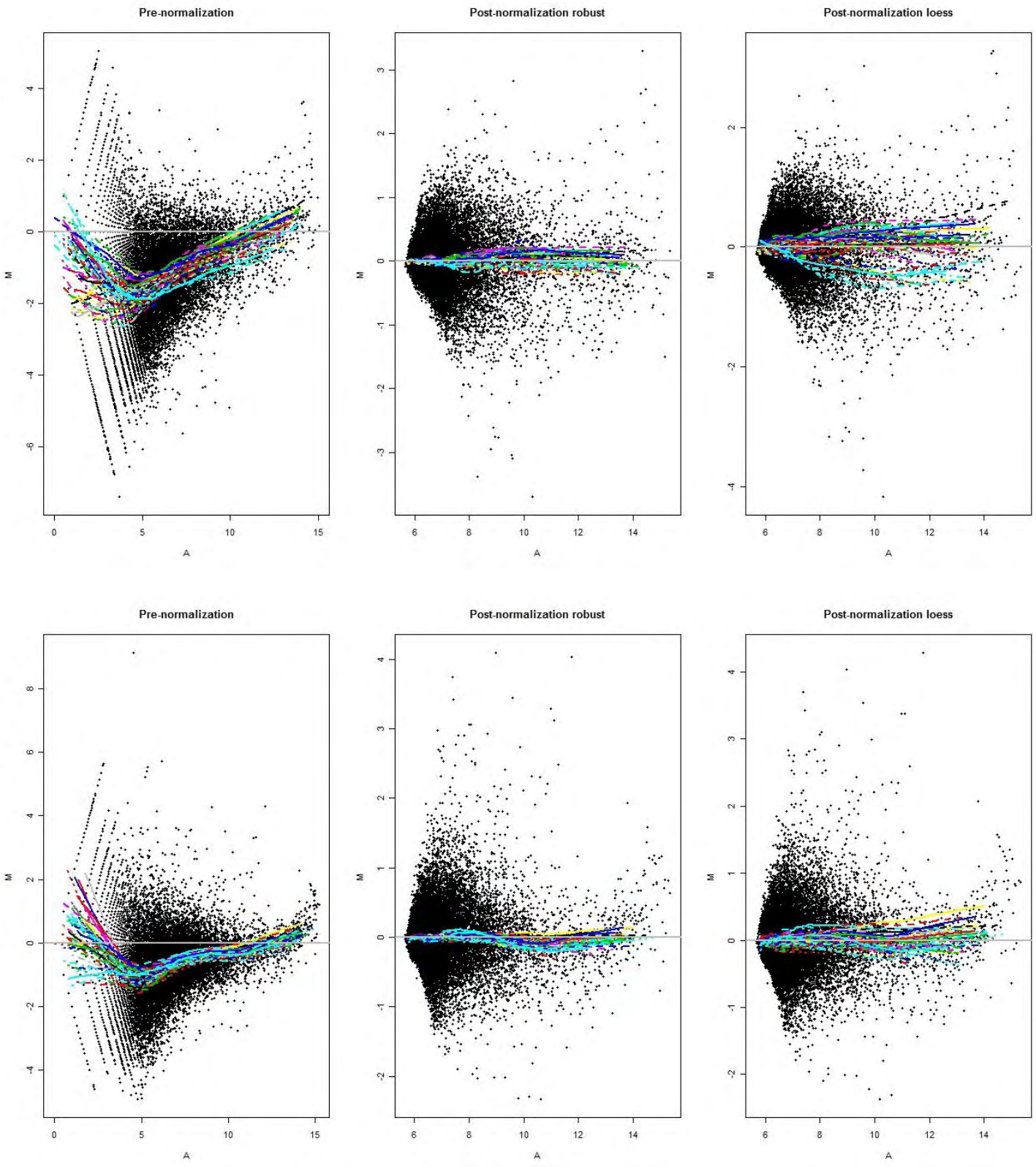


Figure2 - Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 3 and 4, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

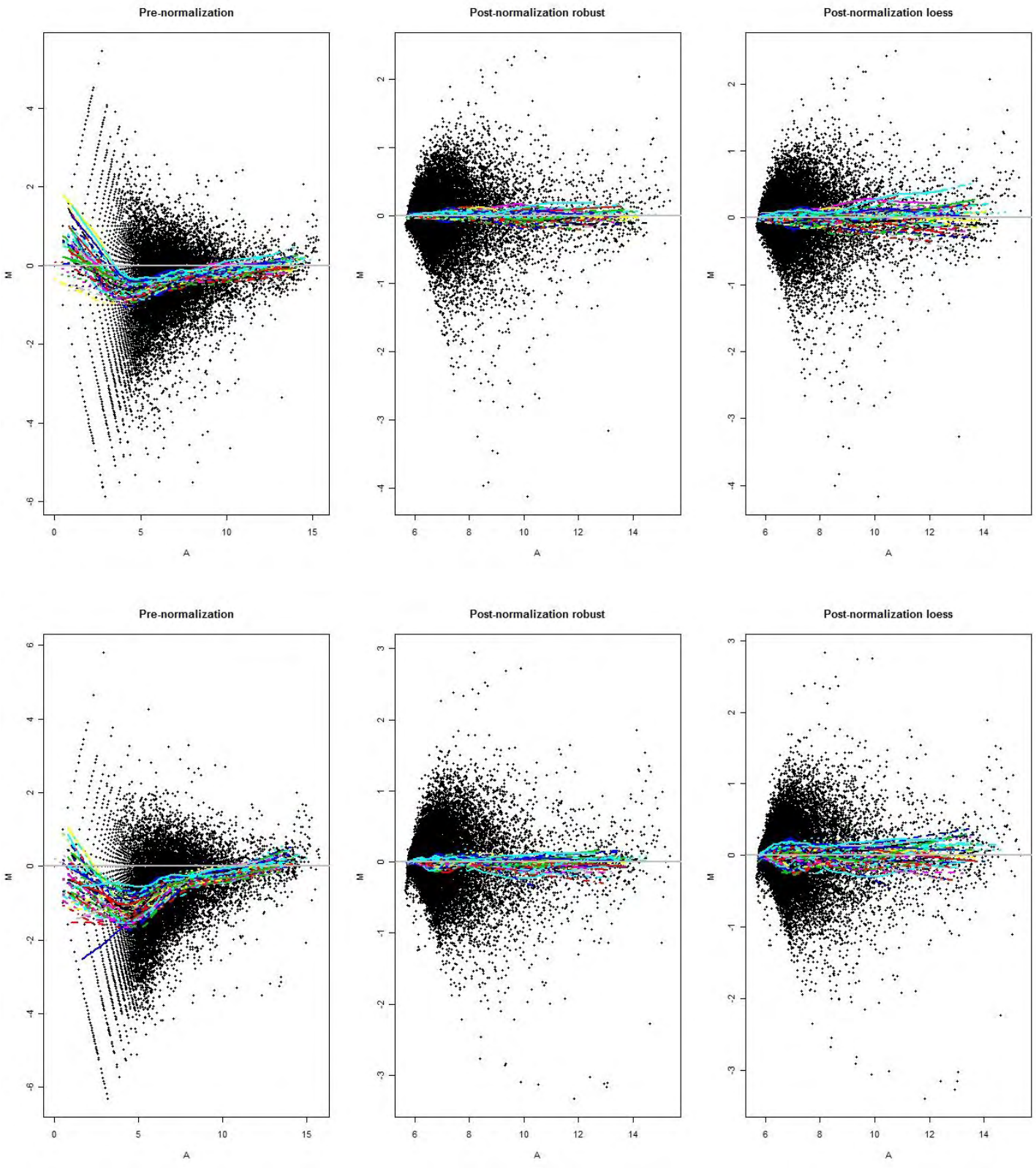


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 5 and 6, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

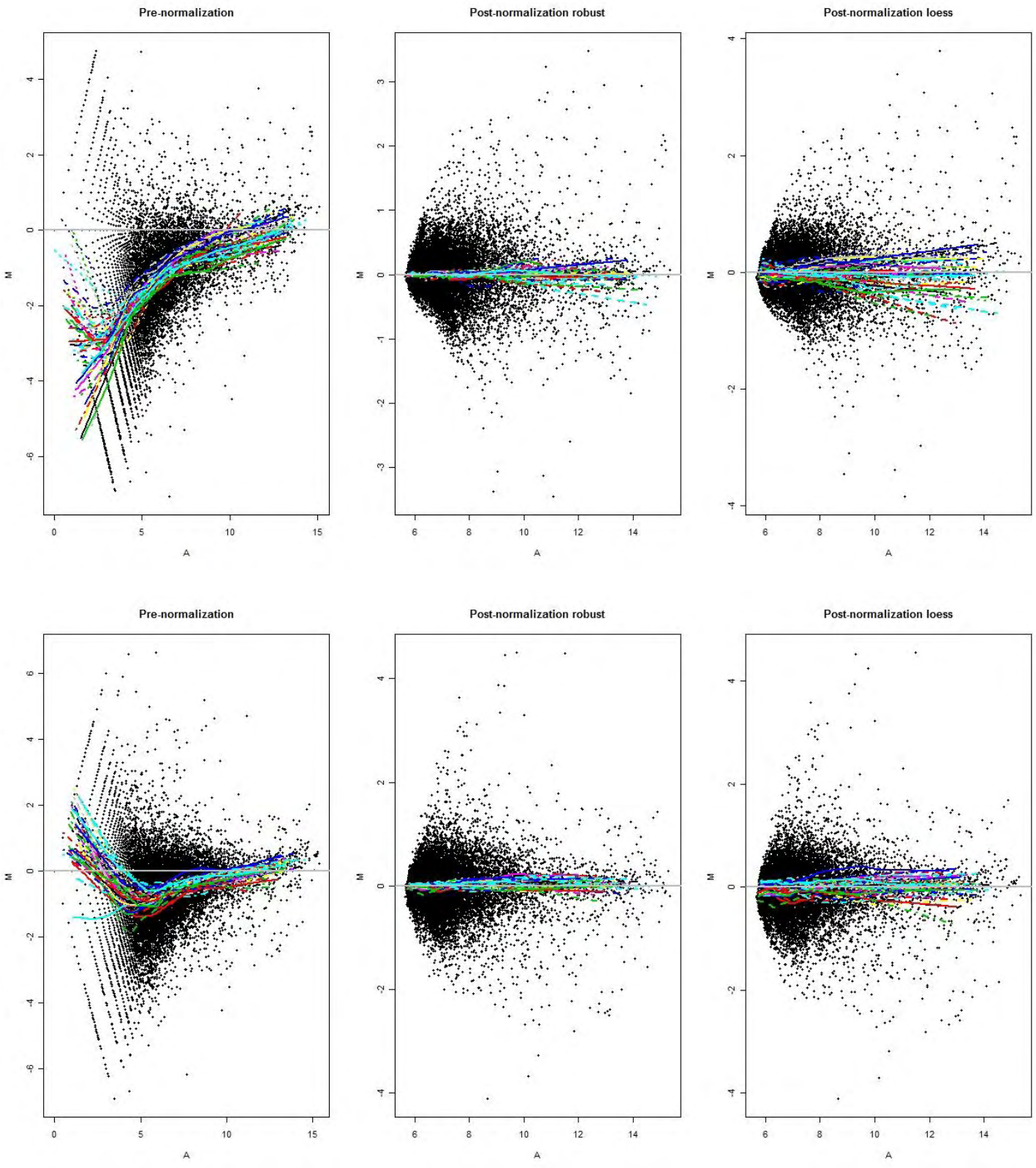


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 7 and 8, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

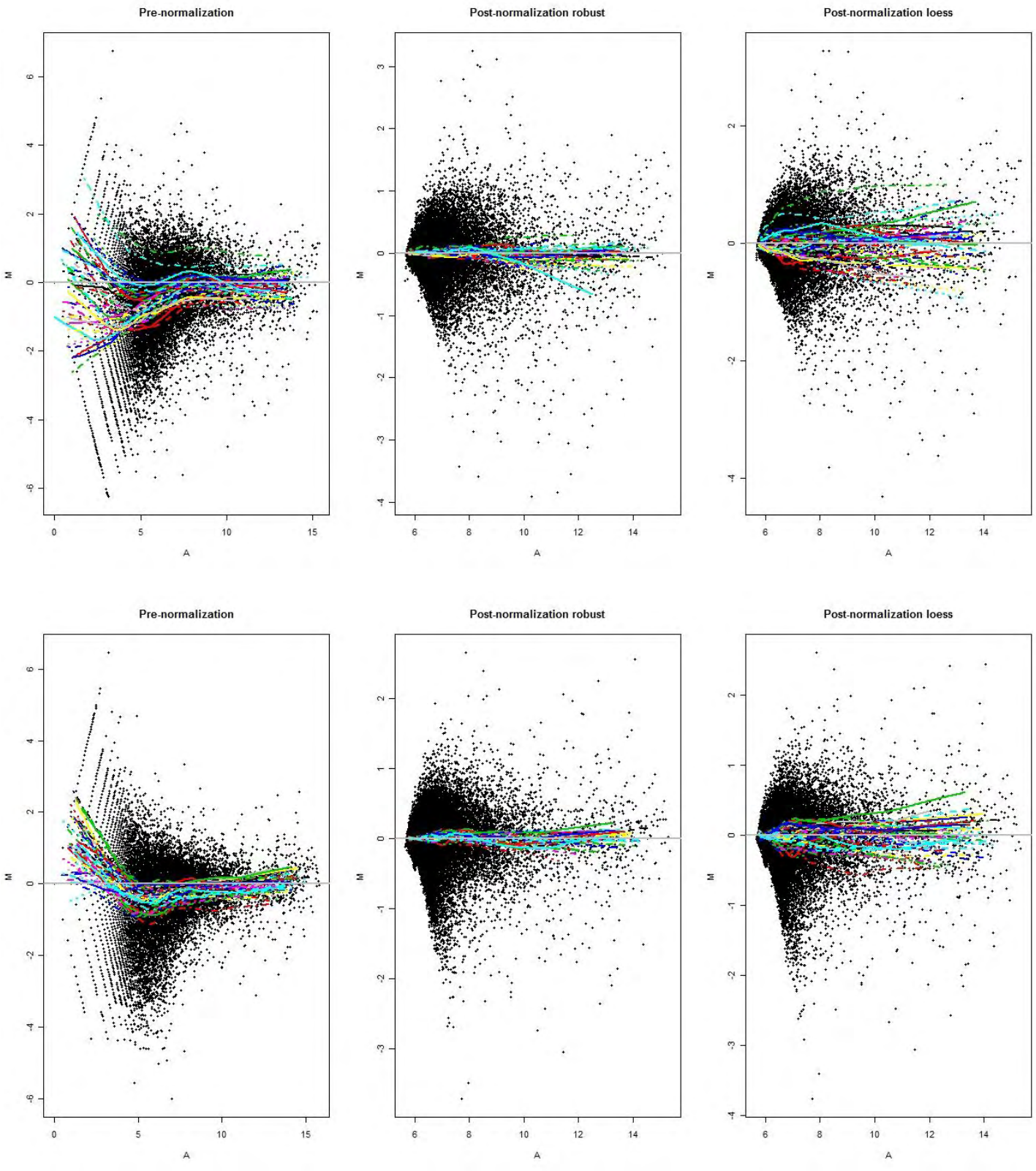


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 9 and 10, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

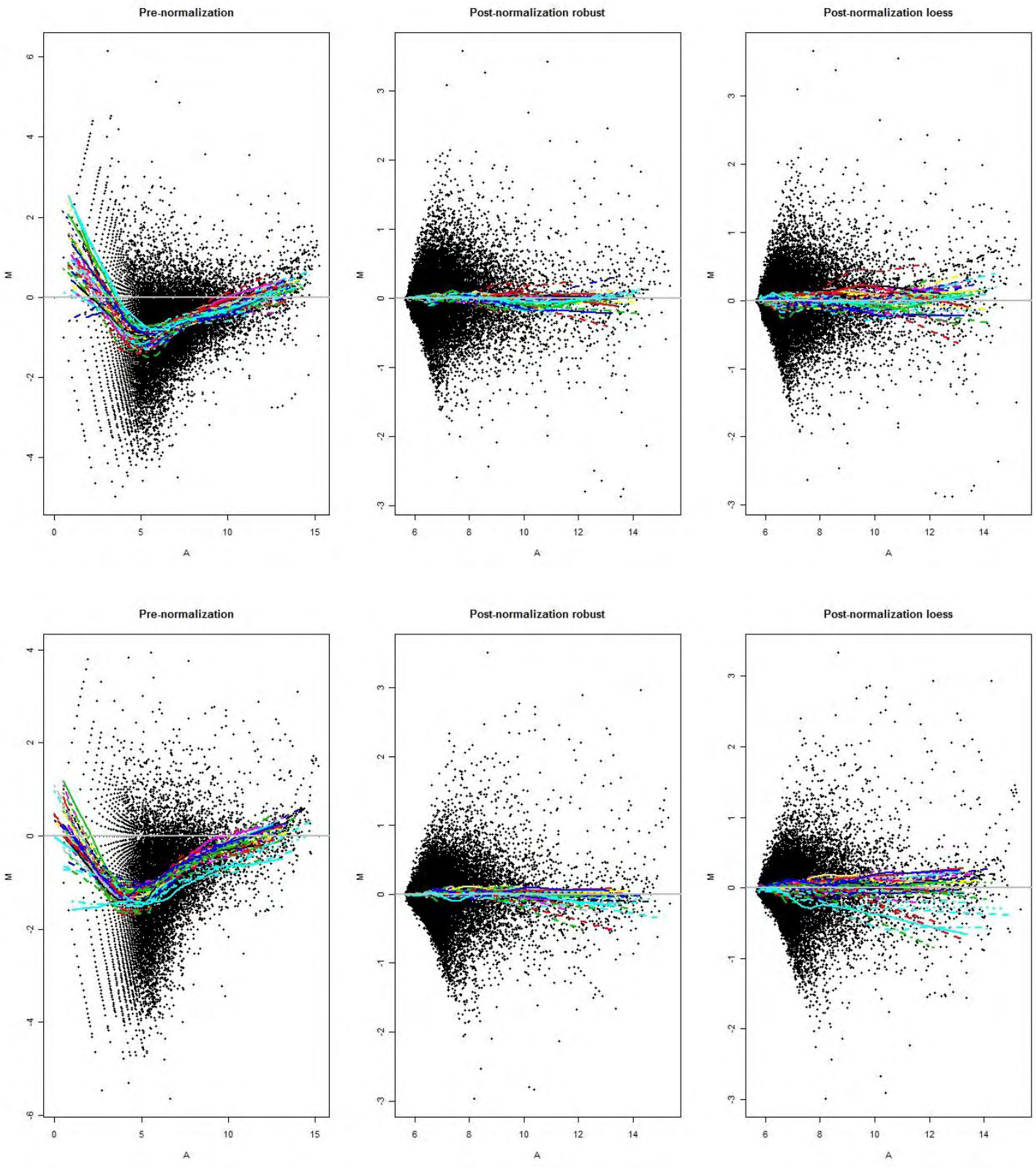


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 11 and 12, respectivelly. Colorful lines represent the trend of the median expression of print-tips groups of each slide.

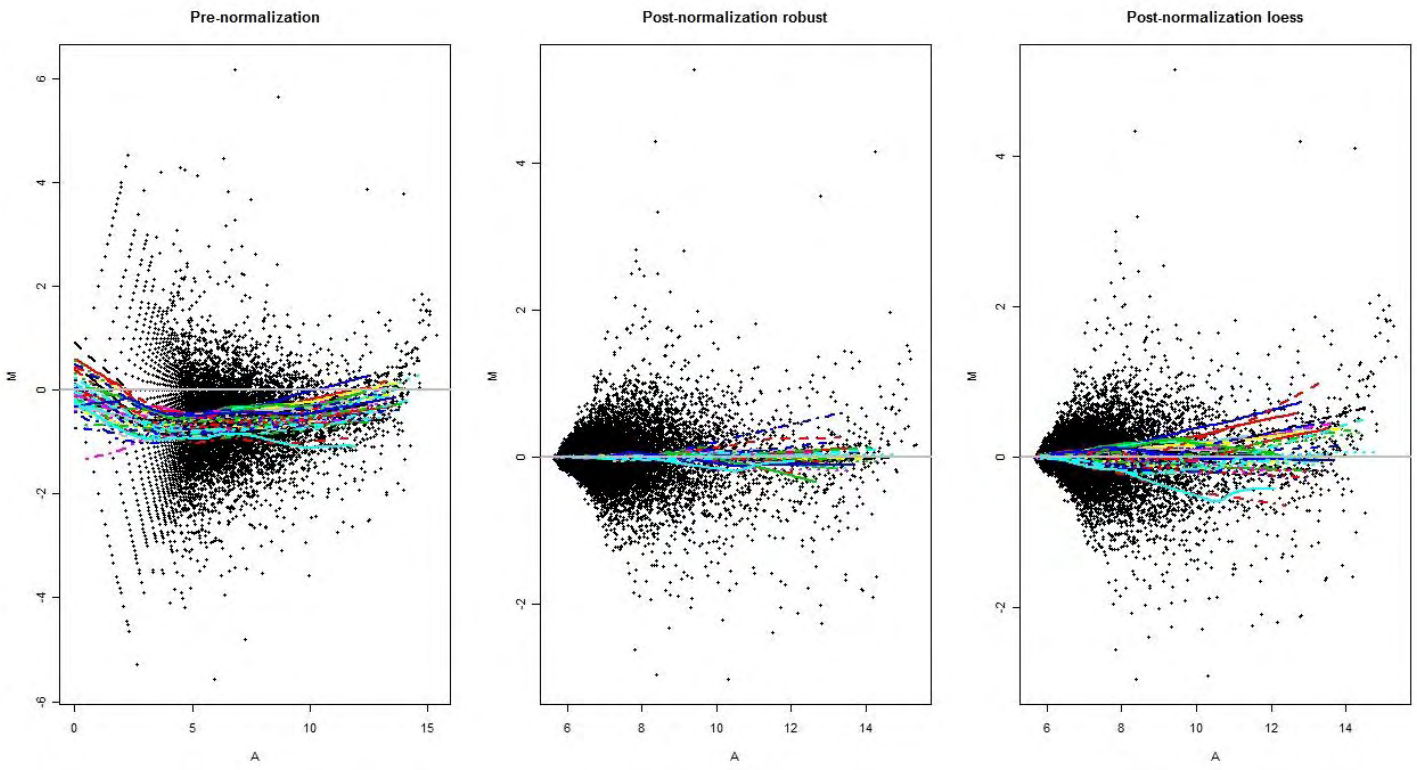


Figure2- Scatherplot of log2 variance ratio versus average log2 mean pre-normalization, post-normalization with Robust Spline method and post-normalization with Loess method of the slide 13. Colorful lines represent the trend of the median expression of print-tips groups of each slide.