Identification of genes affecting intramuscular fat content in Iberian pigs through *Longissimus dorsi* transcriptomic analyses.

<1 line>
M. Muñoz[1,2], J.M. García-Casco[1,2], F. Sánchez-Esquiliche.[3], F. García[2], M.A. Fernández-Barroso[1,2], J.M. Pariente[3], F. Gómez[3], M.C. Rodríguez[2], & L. Silió[2]

[1]Centro de I+D en Cerdo Ibérico INIA-Zafra,,06300 Zafra (Badajoz), Spain.
garcia.juan@inia.es (Corresponding Author)
[2]Departamento de Mejora Genética Animal, INIA, 28040, Madrid, Spain
[3]Sánchez-Romero Carvajal, Carretera San Juan del Puerto s/n, 21290 Jabugo (Huelva), Spain

Summary

Iberian pigs have a high content of intramuscular fat (%IMF) which explains the remarked quality of its meat and dry-cured products. The discovery of genes that codify for proteins involved in intramuscular fat content is very valuable for searching polymorphisms useful in a selection scheme to maintain or improve this trait. The objective of this study was to analyse through RNAseq technique differentially expressed (DE) genes on *Longissimus dorsii* in two groups of a selected Iberian line with extreme breeding values for IMF. A total of 221 and 116 DE genes and potentially new isoforms, respectively, were detected. The Gene Ontology enrichment analyses and Genome-Scale Metabolic Network (Recon) showed an enrichment of DE genes in pathways related with skeletal muscle development, fatty acid metabolism and adipogenesis. In addition to this, the analyses with VarElect showed candidates genes for IMF previously reported as *SCD* and other no previously studied as *EGR1*, *PFKFB3* or *ARID5B* among others. These genes are very promising as candidate genes for intramuscular fat, the detection of polymorphisms and their inclusion in the breeding scheme will be evaluated in a close future.

Keywords: Iberian pigs, intramuscular fat, meat quality, transcriptome

Introduction

Iberian pigs are characterized for having a low growth rate, great appetite and fatness. Its traditional open-air production system includes a finish-fattening period (up to 160 kg approx.) known as *montanera* based on an *ad libitum* intake of acorns and grass (Lopez-Bote, 1998). Both the genetic characteristics of the breed and this production system favor a high content of intramuscular fat (IMF) which plays a major role in the determination of meat quality traits (Wood *et al.*, 2008). This trait is positively correlated with sensorial traits as flavour, juiciness and acceptability of pork (Shi-Zheng and Su-Mei, 2009) and also affects meat tenderness (van Laack *et al.*, 2001).

Nowadays, there are not many selection programs implemented in Iberian pig lines, especially in those fattening under a *montanera* system. However, since 2012, a breeding program focused on the improvement of premium-cuts yield and meat quality of pigs fattened in this extensive system is being developed (Muñoz *et al.*, 2016). The analyses of the transcriptome of individuals divergent for IMF could provide a huge detailed insight of the genes that codify for proteins involved in pathways affecting IMF content. Therefore, the objective of the current study was to investigate the differentially expressed genes in *Longissimus dorsi* between two groups of Iberian pigs with extreme breeding values for

%IMF in order to detect candidate genes which could be eventually incorporated to the selection scheme.

Material and methods

Animal material, diets, RNA isolation and sequencing

The IMF content was measured in loin samples of 914 purebred Iberian castrated males fed with acorn and grass during the fattening period and slaughtered in 15 batches. Breeding values (BVs) for this trait were estimated according to the following mixed model:

$$\mathbf{y = Xb + Za + W_{sm}sm + W_{c}c + e,}$$

where **y** is the %IMF of each animal, **b** represents the systematic effects in which the slaughter weight was fitted as a covariate, **a** is the vector of the additive genetic effects (BVs), **sm** is the vector of the batch random effect, **c**, the vector of the litter random effect and **e,** the vector including the residual effects. **X**, **Z**, $\mathbf{W_{sm}}$ and $\mathbf{W_{c}}$ are the incidence matrices. The results were used to choose the most extreme animals. Twelve non-siblings individuals with extreme BVs for %IMF were selected and arranged in two groups (six HIGH and six LOW). The averaged BVs were 3.07 and -1.38 for the HIGH and the LOW group respectively.

Samples of *Longissimus dorsi* muscle were collected at slaughter and stored at -80ºC. Ribopure High Quality total RNA kit (Ambion, Austin, TX) was used to extract total RNA. The integrity of the RNA was assessed with an Agilent 2100 Bioanalyzer device (Agilent Technologies, Santa Clara, CA). Paired-end libraries were set to be sequenced on an Illumina Hi-Seq 2000 (Fasteris, Plan-les-Ouates, Switzerland) with four samples per lane generating paired-end reads of 75 bp.

FastQC was used to assess the quality of the raw sequencing and *Trim Galore* to trim them. Filtered reads were mapped against the pig reference genome (Sscrofa11.1) using *TopHat* v2.1.0 (Trapnell *et al.*, 2009) through the alignment of the reads first to the ENSEMBL (11.1) transcriptome annotation. Transcripts were assembled using *Cufflinks* v2.2.1 (Trapnell *et al.*, 2012). *Cuffcompare* tool was used to classify the transcripts.

Differential Expression Analysis, Gene Functional Classification and Network Analyses

The estimation of the expression values in FPKMs and differential expression analyses between the HIGH and the LOW groups of the annotated genes and the newly predicted isoforms were carried out using *Cuffdiff*. Genes and new isoforms with a minimum mean group expression of 0.5 FPKMs and a fold change of the expression differences between HIGH and LOW groups of 1.2 were filtered. The R package *q*-value (Storey and Tibshirani, 2003) was employed for taking into account of multiplicity of tests. Those genes and new isoforms with a *p*-value lower than 0.05 and a *q*-value lower than 0.10 were considered as differentially expressed.

Functional analyses of the DE genes were carried out by examining GO enrichment with FatiGO using GO database. In addition, VarElect tool was used in order to identify genes associated with the terms intramuscular fat and fat.

Results and Discussion

Characterization of *longissimus dorsi* transcriptome and differential expression analyses

After trimming and filtering, 1,623.17 million reads were obtained in the 12 *longissimus dorsi* samples. A percentage of 95.5 % of the reads was mapped against the pig reference genome. *Cufflinks* tools revealed a total of 137,022 transcripts expressed in the 12 samples. Although the new Sscrofa11.1 annotation build was employed in the assembling of the transcriptome, a high percentage of transcripts corresponding to new isoforms, falling within an intron or within intergenic regions (67.88%) was identified. This high quantity could be attributed that although Sscrofa 11.1 new version contains an improved version of the genome annotation, this is still incomplete. The distribution of the expression values obtained with *Cuffdiff* reveals about 15,000 out of the 25,878 annotated genes no expressed in *Longissimus dorsi* (Figure 1).

The differential expression analyses of the annotated genes showed a total of 221 DE genes. While 78 of these genes were up-regulated in the HIGH group, the remaining 143 genes showed greater expression in the LOW group. In addition to this, 116 DE potentially new isoforms were detected, 44 showed higher expression in the HIGH group and 72 in the LOW one.

Functional analyses interpretation

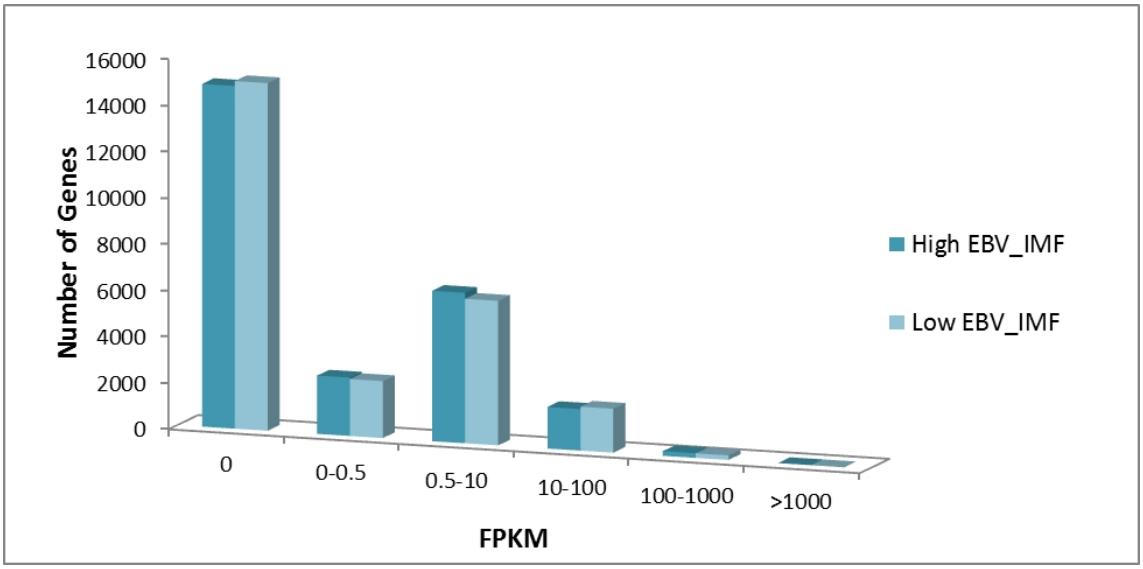
The GO enrichment analyses revealed different biological pathways enriched in DE genes related with skeletal muscle growth, development and differentiation (GO: 0007519, GO: 0060538 and 0035914), fatty acid metabolism as fatty acyl-coA metabolic process (GO: 0035337) or fat deposition as adipose tissue development (GO: 0060612). In addition to this, the Genome-Scale Metabolic Network (Recon) revealed an enrichment of genes up-regulated in the HIGH group in the *fatty acid elongation* pathway pointing out an overstimulation of the lipogenesis in muscle of animals with high BVs for %IMF. These results agree with other studies which report an overexpression of genes related with lipogenesis in pigs with higher IMF content (Cánovas *et al.*, 2010; Hamill *et al.*, 2013).

The GO functional analyses revealed a total of 25 DE genes in pathways related with fatness and skeletal muscle development. In addition, VarElect showed a total of 42 genes with a function associated with IMF. Some examples are *FASN*, *SCD* and *ELOVL6* involved in fatty acid synthesis, *EGR1* which inhibits lipolysis, *PFKFB3*, involved in the glycolysis or *ARID5B* and *DGAT2* involved in adipose tissue development among others.

These genes, some of which have not exhaustively been studied so far, are very promising candidate genes which can harbour polymorphisms associated to IMF and body composition in Iberian pigs. The search of these genetic variants is the next step of this study.

<1 line>

Figure 1. Gene expression distribution of the 25,878 genes annotated in the pig genome in fragments per kilobase of transcript per million mapped fragments (FPKMs) for the HIGH and LOW groups.



Acknowledgment

This study was financially supported by P2013/ABI2913: MEDGAN grant.

List of References

Cánovas A., Quintanilla R., Amills M. and Pena R. N. (2010) Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics,* 11**,**(1): 372.

Hamill R. M., Aslan O., Mullen A. M., O'Doherty J. V., McBryan J., Morris D. G. and Sweeney T. (2013) Transcriptome analysis of porcine M. semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction. *BMC Genomics,* 14**,**(1): 453.

Lopez-Bote C. (1998) Sustained utilization of the Iberian pig breed. *Meat Sci.,* 49: S17-S27.

Muñoz M., Sánchez-Esquiliche F., Caraballo C., Gómez F., Pariente J. M., Silió L., Rodríguez C. and García-Casco J. M. (2016) Animal breeding scheme applied to the quality of pure Iberian montanera pigs. 9th International Symposium on the Mediterranean Pig, Portalegre.

Ros-Freixedes R., Gol S., Pena R. N., Tor M., Ibáñez-Escriche N., Dekkers J. C. and Estany J. (2016) Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs. *PLoS One,* 11**,**(3): e0152496.

Shi-Zheng G. and Su-Mei Z. (2009) Physiology, affecting factors and strategies for control of pig meat intramuscular fat. *Recent patents on food, nutrition & agriculture,* 1**,**(1): 59-74.

Storey J. D. and Tibshirani R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences,* 100**,**(16): 9440-9445.

Trapnell C., Pachter L. and Salzberg S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,**(9): 1105-1111.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D. R., Pimentel H., Salzberg S. L., Rinn J. L. and Pachter L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.,* 7**,**(3): 562-578.

van Laack R. L., Stevens S. G. and Stalder K. J. (2001) The influence of ultimate pH and intramuscular fat content on pork tenderness and tenderization. *J. Anim. Sci.,* 79**,**(2): 392-397.

Wood J. D., Enser M., Fisher A. V., Nute G. R., Sheard P. R., Richardson R. I., Hughes S. I. and Whittington F. M. (2008) Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci,* 78**,**(4): 343-358.