DARLENE ANA SOUZA DUARTE

VALIDATING GENOME ASSOCIATION STUDIES FOR MEAT QUALITY AND CARCASS TRAITS IN PIGS THROUGH GENE NETWORKS AND META-ANALYSIS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para a obtenção do título de *Magister Scientiae.*

VIÇOSA
MINAS GERAIS – BRASIL
2015

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
D812v
2015

Duarte, Darlene Ana Souza, 1990-
Validating genome association studies for meat quality and carcass traits in pigs through gene networks and meta-analysis / Darlene Ana Souza Duarte. – Viçosa, MG, 2015.
viii, 43f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.
Orientador: Fabyano Fonseca e Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Genômica. 2. Genes - Análise. 3. Carne. 4. Carcaça. I. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 572.86

DARLENE ANA SOUZA DUARTE

VALIDATING GENOME ASSOCIATION STUDIES FOR MEAT QUALITY AND CARCASS TRAITS IN PIGS THROUGH GENE NETWORKS AND META-ANALYSIS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para a obtenção do título de *Magister Scientiae.*

APROVADA: 20 de fevereiro de 2015

Profa.Simone E. F. Guimarães
(Coorientador)

Prof. Moysés Nascimento

Prof. Fabyano Fonseca e Silva
(Orientador)

AGRADECIMENTOS

A Deus, por ter abençoado meus caminhos até chegar aqui.

A Universidade Federal de Viçosa, por ter me proporcionado a oportunidade.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais, pela concessão da bolsa.

A meu orientador Fabyano Fonseca e Silva pelo tempo, confiança e paciência durante todos esses anos.

Aos meus coorientadores Simone Eliza Facioni Guimarães e Paulo Sávio Lopes pelos ensinamentos e grande contribuição.

A todos meus amigos do DZO pelas conversas e bons momentos.

Aos meus amigos do LABTEC, especialmente Renatinha e Lucas por terem dedicado horas e horas de seu tempo às minhas análises. A “tia” Vanessa por estar sempre comigo em todos momentos e por me suportar todo esse tempo, a Carolzinha por sempre cuidar de mim e pelos bons conselhos, a meu irmão de coração, André pela amizade e por estar sempre cuidando de mim também, ao Walmir por toda ajuda e paciência.

A todos os amigos do melhoramento pela aprendizagem e risadas.

A minha amiga Cândida pela grande amizade, boas gargalhadas e paciência.

Aos meus pais por todo apoio em todas minhas decisões, pelos inúmeros conselhos e por serem meus exemplos de ser humano.

A minha irmã Daniela por todos ensinamentos e por também ser um exemplo para mim.

BIOGRAFIA

Darlene Ana Souza Duarte, filha de José Maria Duarte e Maria Izabel de Souza Duarte, nascida em 12 de fevereiro de 1990, em Santa Cruz do Escalvado, Minas Gerais, Brasil.

Iniciou o curso de Zootecnia em março de 2008 pela Universidade Federal de Viçosa, graduou-se em maio de 2013.

Em abril de 2013 começou o mestrado em Genética e Melhoramento pela mesma universidade.

SUMMARY

RESUMO

DUARTE, Darlene Ana Souza, M.Sc., Universidade Federal de Viçosa, Fevereiro de 2015. **Validação estudos de associação genômica para qualidade de carne e características de carcaça em suínos através de redes gênicas e meta-análise.** Orientador: Fabyano Fonseca e Silva. Co-orientadores: Simone Eliza Facioni Guimarães e Paulo Sávio Lopes.

Um grande número de locus de características quantitativas (QTLs) para qualidade de carne e carcaça tem sido identificado em diversos estudos, mas a arquitetura genética envolvida nessas características ainda é pouco compreendida. Dessa forma, uma metodologia que permite estudar genes e vias que afetam essas características oferece vantagens e aumenta o conhecimento dos mecanismos fisiológicos. Com esta finalidade, uma metodologia chamada Matriz de Pesos de Associação (AWM) foi utilizado para investigar a base genética dessas características e gerar redes gênicas com base na co-associação de pares de SNPs através dos fenótipos. Foi realizado estudos de associação genômica para 12 características e 144 SNPs foram significativos. Uma meta-análise foi realizada para validar os resultados obtidos no estudo de associação genômica do presente estudo. Alguns SNPs significativos encontrados nos estudos de associação genômica deste trabalho estão perto de QTLs achados nos estudos usados na meta-análise. Portanto, os resultados da meta-análise corroboram os resultados do estudo de associação genômica do presente trabalho. Os SNPs considerados significativos nos estudos de associação genômica foram selecionados para a construção da AWM. Através dessa metodologia, foi possível encontrar 45 genes e estes foram usados para a construção de uma rede com base na correlação entre eles. Além disso, foram identificados 25 fatores de transcrição (FT) fortemente relacionados ($p < 0,001$) com os genes dessa rede. Os três top FT (Sox5, NKX2-5 e T) foram escolhidos para a construção de uma rede com suas vias e ontologia gênica. Os genes da rede e associado com os FT estão envolvidos no metabolismo do tecido adiposo e do músculo esquelético. Nossos resultados sugerem que os genes e FT identificados aqui são importantes no controle de características de qualidade de carne e carcaça. No entanto, esforços devem ser feitos a fim de estudar mais

detalhadamente as novas interações gene-gene aqui identificados, bem como, os fatores de transcrição chave e vias envolvidas nestas características.

ABSTRACT

DUARTE, Darlene Ana Souza, M.Sc., Universidade Federal de Viçosa, February 2015. **Validating Genome Association Studies for Meat Quality and Carcass Traits in Pigs Through Gene Networks and Meta-Analysis.** Adviser: Fabyano Fonseca e Silva. Co-Advisers: Simone Eliza Facioni Guimarães and Paulo Sávio Lopes.

A large number of quantitative trait loci (QTLs) for meat quality and carcass traits has been identified in several studies, but the genetic architecture remains poorly understood. Thus, a methodology that allows study genes and pathways that affect these traits would offers many advantages and increase the knowledge of physiological mechanisms. With this purpose, a methodology named Association Weight Matrix (AWM) was used to investigate the genetic basis of these traits and generate gene network based on the co-association of pair-wise SNPs across phenotypes. We performed genome association studies for 12 traits and 144 SNPs was found to be significant. A meta-analysis was performed to validate the results obtained in the genome association study in this present study. Some significant SNPs found in the genome association studies of this work are close to QTLs findings in the studies from meta-analysis. Therefore the results from meta-analysis corroborated those of the genome association studies of the present work. The significant SNPs from genome association studies were selected to build the AWM. Through this methodology, we could found 45 genes, these genes were used to build a gene network based on pairwise correlations between them. Besides, we identified 25 transcription factors (TF) strongly related (p-value<0.001) with genes in the network. The top three TF (Sox5, Nkx2-5 and T) were choosing for construction of a network with their pathways and gene ontology. The genes from network and associated with this TF were involved in metabolism of adipose tissue and skeletal muscle. Our results suggest

that genes and TF identified here are important in the control of meat quality and carcass traits. However, further efforts should be made in order to study in more detail the new gene-gene interactions here identified, as well as, the key transcription factors and pathways involved in these traits.

Chapter 1

GENERAL INTRODUCTION

During the last ten years, the use of molecular markers, revealing polymorphism at the DNA level, has been playing an increasing part in animal genetics studies (Vignal et al., 2002). Single nucleotide polymorphisms (SNPs) have emerged as genetic markers of choice because of their high-density and relatively even distribution in the genome and have been used by many groups for fine mapping disease loci and for candidate gene association studies (Chen and Sullivan 2003).

Recently molecular genetics has reached great advances due to a development of dense maps of SNP; high-throughput sequencing together with modern genotyping platforms.

Most quantitative traits (like carcass and meat quality) are controlled by several genes and is affect by the environment. The genome association studies use high-throughput genotyping technologies to assay hundreds of thousands of single-nucleotide polymorphisms that can be close to these genes make possible to relate them to clinical conditions and measurable traits through genome association studies (Pearson and Manolio, 2008).

A systems biology approach for the genetic dissection of complex traits has been described to analyze the results from genome association studies and combiner a series of phenotypes of interest. Such a strategy would be particularly useful when investigating the genetic basis of complex traits that, by definition, are influenced by numerous genes and pathways (Reverter and Fortes, 2013a). This methodology named Association Weight Matrix (AWM) is used to exploit the results from genome association studies and, in combination with network inference algorithms, generate gene networks with regulatory and functional significance (Fortes et al., 2010). Thus, we can get more information about genes that are controlled traits of interest.

Single nucleotide polymorphism (SNP)

As the Human Genome Project (HGP) progressed and the nucleotide sequence of the human genome was being unveiled, an evident finding was the large number of point variations found in comparison of corresponding segments of the genome. Even before the first draft of the complete sequence become available, a large portion of the scientific community has turned attention to these small and abundant variations scattered throughout the genetic code.

These variations, which are the most frequent type found in DNA, are called SNPs and they are valuable markers for high-throughput genetic mapping, genetic variation studies and association mapping. SNPs are single base pair positions in genomic DNA at which a change of alternative alleles occurs between members of the same specie or paired chromosomes in an individual. When the mutation presents an abundance of the least frequent allele of 1% or greater in the population evaluated it is referred to as a single nucleotide polymorphism (Brookes, 1999). SNP markers could be bi-, tri- or tetra-allelic polymorphisms. Although, most of the SNP markers are bi-allelic. It is rare to find tri- or tetra-allelic, because of the low probability of two independent base changes occur at a single position (Vignal et al., 2002)

SNP markers belong to the last-generation molecular markers, occur at high frequencies in both animal and plant genomes. In the human genome, about 90% of total polymorphisms are differences in single base of DNA (Collins et al., 1998). The average overall frequency of SNPs is estimated one per 1000 bp or less (Weiner & Hudson, 2002). In the pig genome it is estimated one SNP per 609 bp (Fahrenkrug et al., 2002). These randomly occurring changes are passed from generation to generation and account for a high proportion of the DNA differences between individuals. According to the dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP), there are more than 9 million SNPs distributed along the human genome.

In general, SNPs can be found much less frequently in coding regions of the genome than in noncoding regions (Caetano, 2009). SNPs in noncoding regions, whilst they do not alter encoded proteins, serve as important genetic or physical markers for comparative or evolutionary genomics studies (Kim & Misra, 2007). The most frequent substitutions that occur in DNA are the transitions, which consists in exchanges between two purines (A/G or G/A) or

two pyrimidines (C/T or T/C). Less frequent, transversions are substitutions of a purine for a pyrimidine or vice versa (Brookes, 1999).

The development of SNP markers allows automatizing and enhancing the effectiveness of genotype analysis. Thus, it is possible genotyping hundreds of thousands individuals for several thousand SNP markers in a few hours (Williams, 2005). The high-density SNP-Chips enabled the generation of new applications that will bring significant advances for animal breeding programs.

Genome Association Studies

A Genome Association Studies is an approach that involves rapidly scanning markers across genome to identify genetic associations with observable traits. Biotechnological advances in the automation of genotyping process, which started to be done on a large scale, allowed the development of new classes of markers, among which it is possible to highlight the SNPs.

In the genome-wide association studies (GWAS), a dense set of SNPs across the genome is genotyped to survey the most common genetic variation for common diseases and other phenotypes of interest (Hirschhorn and Daly, 2005). The genotyping technology being considerably improved and becoming cheaper in recent years, has enabled the implementation of GWAS.

The typical GWA study can be conducted following these steps: (1) selection of a large number of individuals with the trait of interest; (2) DNA isolation, genotyping, and data review to ensure high genotyping quality and (3) statistical tests for associations between the SNPs passing quality thresholds and the trait (Pearson and Manolio, 2008).

Thus, it is possible through identification of variation, find candidate genes influencing the traits under study.

Meta-analysis

Meta-analysis consists of statistical methods for combining results of independent studies. One aim of combining results is to obtain increased power – studies with small sample size are less likely to find effects even when they exist (Guerra and Goldstein, 2009).

The genetic effect sizes are almost invariably moderate to small in magnitude and single studies, even if large, are underpowered to detect them with confidence. Meta-analysis of any genome-wide association studies improve the power to detect more association, and investigate the consistency or heterogeneity of these associations across diverse datasets and study populations (Zeggini and Ioannidis, 2009).

To perform meta-analysis is required careful planning and execution of the following six general steps: formulate a specific purpose and explicitly define the outcome to be extracted from each study (e.g. effect size, p-value, etc); identify relevant primary studies; establish inclusion/exclusion criteria for primary studies; detail data abstraction and acquisition; decide on data analysis methods, including a careful investigation of data quality and between-study heterogeneity and interpret the results, and decide on appropriate follow up steps (Guerra and Goldstein, 2009).

Meta-analysis of GWAS should be viewed partly as an opportunity to replicate candidate variants using formal statistical tests and partly as an opportunity to discover new variants or to explore the way that variants operate (Thompson *et al.*, 2011).

The method used in the present study is the inverse variance method. This method is equivalent to combining likelihoods coming from separate studies, using a quadratic approximation. Denote coefficients of regression estimated in N studies as βi, and associated squared standard errors of the estimates as s_i^2 where i ∈ 1,2,...,N. Define weights for individual studies as

$$w_i = \frac{1}{s_i^2}$$

Then the pooled estimate of the regression coefficient is

The weights have straightforward interpretation: the bigger the weight of the study (meaning the small is the standard error in the study), the larger is the contribution from this study onto the pooled estimate. The standard error of the pooled estimate is computed as

$$s^2 = \frac{1}{\sum_{i=1}^{N} w_i}$$

and the x^2-test for association is computed in standard manner as

$$T^2 = \frac{\beta^2}{s^2} = \frac{(\sum_{i=1}^{N} w_i \beta_i)^2}{\sum_{i=1}^{N} w_i}$$

Association Weight Matrix (AWM)

Association Weight Matrix (AWM) is a novel procedure to exploit the results from genome-wide association studies (GWAS) and, in combination with network inference algorithms, generate gene networks with regulatory and functional significance. AWM was created with the objective of treat the results from GWAS beyond a simple enumeration of the association of (potentially very many) genetic markers (Reverter and Fortes, 2013a).

The principal purpose of AWM is the generation of gene networks based on the co-association of pair-wise SNPs across phenotypes. The basic metric measuring the strength of such co-association is the correlation coefficient calculated from standardized SNPs additive effects.

The condition to build the AWM is the availability of the results from GWAS of a series of phenotypes. For convenience, it is assumed that standard quality control filters (such as the removal of SNP with low minor allele frequency, with poor genotype calling rate or in gross violation of the Hardy–Weinberg equilibrium) have been applied prior to obtaining the GWAS results (Reverter and Fortes, 2013a). To build the AWM three information are required: 1) the estimated additive effects of the SNP to the phenotypes, 2) the p-values associated with the significance of those associations, and 3) the mapping of the SNP to the genome, including the distance to the nearest annotated gene. In addition, one of the phenotypes under scrutiny should be identified as the key phenotype, considering its greater focus within the overall study (Reverter and Fortes, 2013b).

The advantage of using the AWM methodology is in the fact that AWM handle with more than one trait at the same time (taking advantage of the correlation between them). Thus, increase the power to detect genes that act

directly or indirectly on a key trait, which would not be possible with the GWAS single-trait that focuses on the most significant SNPs for one trait at a time.

Objectives

The aim was to validate the results of genome association studies by meta-analysis and validated the gene network by identification of transcriptional factors related with those genes. In addition, Find candidate genes for carcass and meat quality traits and, based in AWM, infer about interaction between these genes.

References

Brookes, AJ 1999. The essence of SNPs. Gene, 234, p.177-186.

Caetano, AR 2009. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. R. Bras. Zootec., v.38, p.64-71 (supl. especial).

Chen, X and Sullivan, PF 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. The Pharmacogenomics J, 3(2), 77–96.

Collins, FS, Brooks, LD, Chakravarti, A 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Research, p.1229-1231.

Fahrenkrug, SC, Freking, BA, Smith, TPL, Rohrer, GA, Keele, JW 2002. Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. Anim. Genet., 33, p.186-195.

Fortes MRS, Reverter A, Zhanga Y, Collisa E, Nagaraj SH, Jonssona NN, Prayaga KC, Barrisa W and Hawkena RJ 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. PNAS 107, 13642–13647.

Guerra R and Goldstein DR 2009. Meta-analysis and Combining Information in Genetics and Genomics CRC Press.

Hirschhorn JN and Daly MJ 2005. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6, 95-108.

Kim, S and Misra, A 2007. SNP genotyping: technologies and biomedical applications. Annual review of biomedical engineering, 9, p. 289-320.

Pearson TA and Manolio TA 2008. How to interpret a genome-wide association study. JAMA 299, 1335-1344.

Reverter A and Fortes MR 2013a. Association weight matrix: a network-based approach towards functional genome-wide association studies. Methods Mol Biol 1019, 437-447.

Reverter A and Fortes MRS 2013b. BREEDING AND GENETICS SYMPOSIUM: Building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genomewide association studies1,2. J. Anim. Sci. 91, 530-536.

Thompson JR, Attia J and Minelli C 2011. The meta-analysis of genome-wide association studies. Brief Bioinform 12, 259-269.

Vignal, A, Milan, D, SanCristobal, M, Eggen, A 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol., 34, p. 275-305.

Weiner, MP and Hudson, TJ 2002. Introduction to SNPs: discovery of markers for disease. BioTechniques, Suppl(43), p.4-7, 10, 12-3.

Williams, JL 2005. The use of marker-assisted selection in animal breeding and biotechnology. Revue scientifique et technique (International Office of Epizootics), 24, p.379-91.

Zeggini E and Ioannidis JPA 2009. Meta-analysis in genome-wide association studies. Pharmacogenomics 10(2), 191-201.

Chapter 2

Validating Genome Association Studies for Meat Quality and Carcass Traits in Pigs Through Gene Networks and Meta-Analysis[1]

ABSTRACT: A large number of quantitative trait loci (QTLs) for meat quality and carcass traits has been identified in several studies, but the genetic architecture remains poorly understood. Thus, a methodology that allows study genes and pathways that affect these traits would offers many advantages and increase the knowledge of physiological mechanisms. With this purpose, a methodology named Association Weight Matrix (AWM) was used to investigate the genetic basis of these traits and generate gene network based on the co-association of pair-wise SNPs across phenotypes. We performed genome association studies for 12 traits and 144 SNPs was found to be significant. A meta-analysis was performed to validate the present study. All studies used in meta-analysis was found to be significant, some SNPs found significant in this work is closer to QTLs from meta-analysis studies. Therefore the results from meta-analysis corroborated those of the present study. The significant SNPs from genome association studies were selected to build the AWM. Through this methodology, we could found 45 genes, these genes were used to build a gene network based on pairwise correlations between them. Besides, we identified 25 transcription factors (TF) strongly related (p-value<0.001) with genes in the network. The top three TF (Sox5, Nkx2-5 and T) were choosing for construction of a network with their pathways and gene ontology. The genes from network and associated with this TF were involved in metabolism of adipose tissue and skeletal muscle. Our results suggest that genes and TF identified here are important in the control of meat quality and carcass traits. However, further efforts should be made in order

to study in more detail the new gene-gene interactions here identified, as well as, the key transcription factors and pathways involved in these traits.

Keywords: genome association studies, gene network, meta-analysis, pig

---

Article written in the Animal journal format

Introduction

Meat quality and carcass are traits of great relevance in pig production. These traits are influenced by a large number of factors including muscle characteristics (fiber size and type, fat and connective tissue), production and environmental conditions (growth rate, nutrition, age, slaughter conditions) and the genetics of the animals (breed and growth potential) (Davoli and Braglia, 2003).

Given the great importance that intramuscular fat content (IMF) has on meat quality, such as juiciness and flavor, and of its acceptability by consumers and consequently on its economic value (Fernandez et al., 1999; Gao & Zhao 2009; Hocquette et al., 2010), IMF was chosen as a key trait in this study.

A large number of quantitative trait loci (QTLs) for these traits has been identified in several studies, hence it seems appropriate to combine the results from some studies using the meta-analysis approach, thus validate the results obtained from genome association studies through meta-analysis.

Although several QTLs has already been found for meat quality and carcass traits (Rohrer et al., 2005; Duthie et al., 2008 ; Sato et al., 2003), the genetic architecture remains poorly understood. Studying gene interactions and pathways that affect these traits, and specially IMF, increase the knowledge of causative physiological mechanisms. The power of single trait GWAS can be enhanced when considering simultaneously multiple phenotypes because complex traits generally have multiple correlated traits (Puig-Oliveras et al., 2014). With this purpose, systems biology approach for the genetic dissection of complex traits integrating the information of GWAS with network inference

algorithms, named Association Weight Matrix (AWM), was developed (Fortes et al., 2010).

The focus of this work is to identify key regulators and gene interactions determining meat quality and carcass traits in order to improve our knowledge about the architecture of these complex traits. Moreover, to validate the results of genome association studies through meta-analysis approach.

Material and Methods

Experimental population and phenotypic data

The use of animals was reviewed and approved by the Animal Care and Use Committee of the Department of Animal Science of the Universidade Federal de Viçosa, Brazil.

The phenotypic data was obtained from the Pig Breeding Farm of the Department of Animal Science, Universidade Federal de Viçosa (UFV), MG, Brazil. A three-generation resource population was created and managed as described by Band *et al.*, 2005a, Band *et al.*, 2005b. Briefly, two local breed Piau grandsires were crossed with 18 grand dams from a commercial line composed of Large White, Landrace and Pietrain breeds, to produce the F1 generation from which 11 F1 sires and 54 F1 dams were selected. These F1 individuals were crossed to produce the F2 population, of which 345 animals were phenotyped for several meat quality and carcass traits.

Carcass and meat quality traits measured were: hot carcass weight including feet and head (CW), carcass yield including feet and head (CY), carcass length by the Brazilian carcass classification method (CLBRA), carcass length by the American carcass classification method (CLUSA), right half carcass weight

(RHCW), midline thinnest backfat thickness above the last lumbar vertebrae (LBF), the thickest backfat thickness on the shoulder region (SBF), midline backfat thickness immediately after the last rib (LRBF), midline backfat thickeness between last and penultimate lumbar vertebrae (PBF), backfat thickness at P2 site (last rib, 6.5 cm from the midline) (P2BF), bacon depth (BD), loin depth (LD), loin eye area (LEA), total ham weight (THW), trimmed ham weight (TRIMHW), total Boston shoulder weight (TBSW), rib weight (RW), trimmed Boston shoulder weight (TRIMBSW), total picnic shoulder weight (TPSW), trimmed picnic shoulder weight (TRIMPSW), total loin weight (TLW), boneless loin weight (LW), bacon weight (BCW), jowl weight (JW), sirloin weight (SLW), abdominal fat weight (AF), pH 45 minutes after slaughter (pH45), pH 24 hours after slaughter (pH24), lightness (L), yellowness (B), redness (A), saturation (C), intramuscular fat (IMF), drip loss (DL), cooking loss (CL), shear force (SF), total loss (TL) and hue angle (H).

DNA extraction, genotyping and SNP selection

DNA was extracted at the Animal Biotechnology Lab from Animal Science Department of Universidade Federal de Viçosa. Genomic DNA was extracted from white cells of parental, F1 and F2 animals, more details can be found in Band et, al., 2005. The low-density (Habier *et al.*, 2009) customized SNPChip with 384 markers used was based on the Illumina Porcine SNP60 BeadChip (San Diego, CA, USA, Ramos *et al.*, 2009). These SNPs were selected according to QTL positions previously identified on this population using meta-analyses (Silva *et al.*, 2011) and fine mapping (Hidalgo *et al.*, 2013, Verardo *et al.*, 2015). Thus, although a small number of markers have been used, the customized SNPchip

based on previous identified QTL positions ensures an appropriate coverage of the relevant genome regions in this population. From these, 66 SNPs were discarded due to a lack of amplification, and from the remaining 318 SNPs, 81 were discarded due to a minor allele frequency (MAF) < 0.05. Thus, 237 SNPs markers were distributed as follows: SSC1 (56), SSC4 (54), SSC7 (59), SSC8 (30), SSC17 (25) and SSCX (13), being the average distance within each chromosome, respectively: 5.17, 2.37, 2.25, 3.93, 2.68 and 11.0 Mb.

Principal Components

In a study evaluating a large number of traits as performed in this work problems during analysis may occur because correlation of evaluated traits and/or due to a linear combination among them. This problem can be solved by reducing the number of traits eliminating those that contribute less. Thus, we can make use of the Principal Component Analysis (PCA) to obtain independent and non-redundant traits.

PCA is a multivariate technique that transforms a number of correlated quantitative variables into fewer uncorrelated (orthogonal) new variables termed Principal Components (PC). The percentage of the variation of the original traits explained by each PC is equal to the associated eigenvalue, and the weights of the traits in each PC are the terms in the associated eigenvectors (Stearns *et al.*, 2005).

The criterion of the number of variables to be excluded was based on Jolliffe (1972) and Jolliffe (1973). This criterion establishes that the number of discarded variables must be equal to the components whose variance (eigenvalue) is less than 0.7. Thus, traits that presented higher values in the PC with eigenvalue less than 0.7 were excluded from further analysis. The analysis

was performed using the function *prcomp* from R software (http://www.R-project.org/).

Genome Association Studies

Genome association studies were performed for traits that remained after PCA. The effect of each SNP was estimated using the following mixed model:

$$Y = X\beta + Wm + Zu + e,$$

where yi is the phenotypic observation of animal i; X is the incidence matrix relating fixed effects in β with observation in yij , the fixed effects were the general mean; sex, batch and halothane gene genotype; W is the incidence matrix relating the additive association of the kth SNP in m on the jth phenotype; Z is the incidence matrix relating random additive polygenic effects in u with observation in yij, $u \sim N(0, A\sigma_u^2)$ and e_i is the residual term $e_i \sim N(0, \sigma_e^2)$, A is the numerator of the pedigree-based relationship matrix. The polygenic effect was used to point out for possible sub-population structure (e.g. family).

In this context, the significance of each SNP was accessed by t Student test using the FDR criterion to correct for multiple test. The mentioned model was implemented using the lmepack.batch function of the GWAF package from R software.

Meta-analysis

The meta-analysis were performed in order to validate the results found in the genome association studies. In the studies used in meta-analysis we look for

QTLs in the same regions or close to significant SNPs found in our study, since it was not possible to obtain the same markers in all studies. For this purpose 29 studies that found QTLs or SNPs for IMF in the same chromosomes used in the genome association studies of this present study. This study were chosen in Pig QTL database (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>). After choosing these studies, the estimate of regression coefficient and the standard error of the estimate (or, equivalently, the p - value or the test statistics value for association) were used in the meta-analysis.

The meta-analysis were conducted using the *forestplot* function of the *MetABEL* package (Struchalin and Aulchenko, 2014) from R software, which uses the inverse variance method. This method is equivalent to combining likelihoods coming from separate studies, using a quadratic approximation. This method is describe on GenABEL tutorial available at http://www.genabel.org/tutorials.

Thus, from the meta-analysis we can verify whether the QTL or SNP (depending on the study) found are really associated with IMF, and so, these studies can be used to validate the positions of significant SNPs found in the genome association studies of the present work.

Association Weight Matrix (AWM)

The AWM has been used to validate GWAS since reveals new genes and improves understanding of the regulatory mechanism of complex traits. Briefly, the AWM methodology involves selecting SNP from GWAS to generate a matrix

with as many columns as traits and as many rows as selected SNP (Fortes *et al.*, 2010).

In summary, four steps have been used to select SNPs for inclusion in the AWM:

1 - SNPs were ranked according to the number of traits associated ($P < 0.05$) with them, regardless of genomic position. From this rank, the top SNP (which were associated with at least 16% of the traits, were included in the AWM.

2 - We classified each remaining SNP as close, far, very far, or unmapped according to its mapped distance from the nearest annotated gene (Sscrofa10.2 assembly, ftp://ftp.ncbi.nih.gov/genomes/MapView/Sus_scrofa/sequence/). The SNP considered close were located at ≤2.5 kb from the transcription start site of a known gene (either 5′ or 3′).

3 - The far or unmapped (> 1.5 Mb) SNP were discarded from further analysis. Because IMF was the main phenotype of interest, all close SNPs with $P < 0.05$ for IMF were selected. Then, we set to μ the average number of traits associated to SNP that were significant ($P < 0.05$) for IMF and added to AWM all SNP that were significant for ≥μ traits. This step reflects that IMF was the main trait for this study, while retaining information from the other traits as well.

4 - When genes were represented by more than 1 SNP, the SNP with a $P < 0.05$ in the largest number of traits was chosen to represent that gene. If still >1 SNP represented the same gene, selection was based on the smaller sum of P-values across traits. This fourth criterion has the purpose of establishing the mapping rule: 1 SNP = 1 gene. This step keeps linkage disequilibrium from dominating and potentially distorting the AWM results.

The AWM contains as many rows as selected SNP and as many columns as traits. Cell values in the AWM correspond to the z-score normalized effect of the given SNP in that row on the corresponding trait in that column.

Correlated genes in AWM form the basis for gene networks. In these networks genes are nodes connected by edges representing significant SNP correlations. The statistical test used to identify the significant correlations between genes was the partial correlation and information theory (PCIT) algorithm (Reverter and Chan, 2008).

Network and Transcription Factors Analysis

The application Cytoscape (www.cytoscape.org/) was used to build the network from AWM genes.

Providing evidence for the interaction between the Transcription Factors (TF) and its predicted targets via regulatory sequence analysis it is possible to use a kind of *in silico* validation for the TF–target interactions in the SNP genes network (Fortes *et al.*, 2010). For this, we used the software TFM-Explorer (http://bioinfo.lifl.fr/TFM/TFME/), which takes a set of gene sequences, and searches for locally overrepresented transcription factor binding sites (TFBS) using weight matrices from JASPAR (http://jaspar.binf.ku.dk/) data base to detect all potential TFBS, and extracts significant clusters (region of the input sequences associated with a factor) by calculating a score function. This score threshold is chosen to give a p-value equal or less than 10−3 for each position in each sequence, such as described in Touzet and Varre (2007). The top TF related (*p*-value<0.001) were identified and for the three most represented (according to *p*-

values), we construct a network with their interactions (TF-target) and the gene ontology using Cytoscape and GeneCards® tools (http://www.genecards.org/).

RESULTS

Principal Component and Genome Association Studies

A Principal Component Analysis (PCA) was applied to the original traits aiming to avoid redundancy between them. From 38 traits, 26 showed eigenvalue lower than 0.7 indicating that they were redundant (Data supplements Table 1). Thus, 12 traits were retained and considered as the main traits to be used in the association analysis. A total of 144 SNPs were found as significant for these traits being 30, 15, 44, 5, 19, 23, 21, 21, 21, 12, 26, 19 respectively for redness (A), saturation (C), midline thinnest backfat thickness above the last lumbar vertebrae (LBF), intramuscular fat content (IMF), jowl weight (JW), bacon weight (BCW), total loin weight (TLW), trimmed Boston shoulder weight (TRIMBSW), cooking loss (CL), drip loss (DL), total ham weight (THW) and carcass yield including feet and head (CY) (Data supplements Table 2).

Genome association studies were accomplished for those traits and 144 SNPs were found to be significant. The estimated SNPs additive effects were standardized (z-scores) by subtracting the mean and dividing by the phenotype-specific standard deviation.

Meta-analysis

A meta-analysis was performed in order to found QTLs in the same regions or close to significant SNPs found in the present study. Eight studies with a total of 29 QTLs for intramuscular fat were used in this meta-analysis, and all of them

were found to be significant ($p=3.8x10^{-3}$) from pooled studies (Fig.1). It means that the results from the meta-analysis studies is consistent and can be considered to validating the results from the present study.

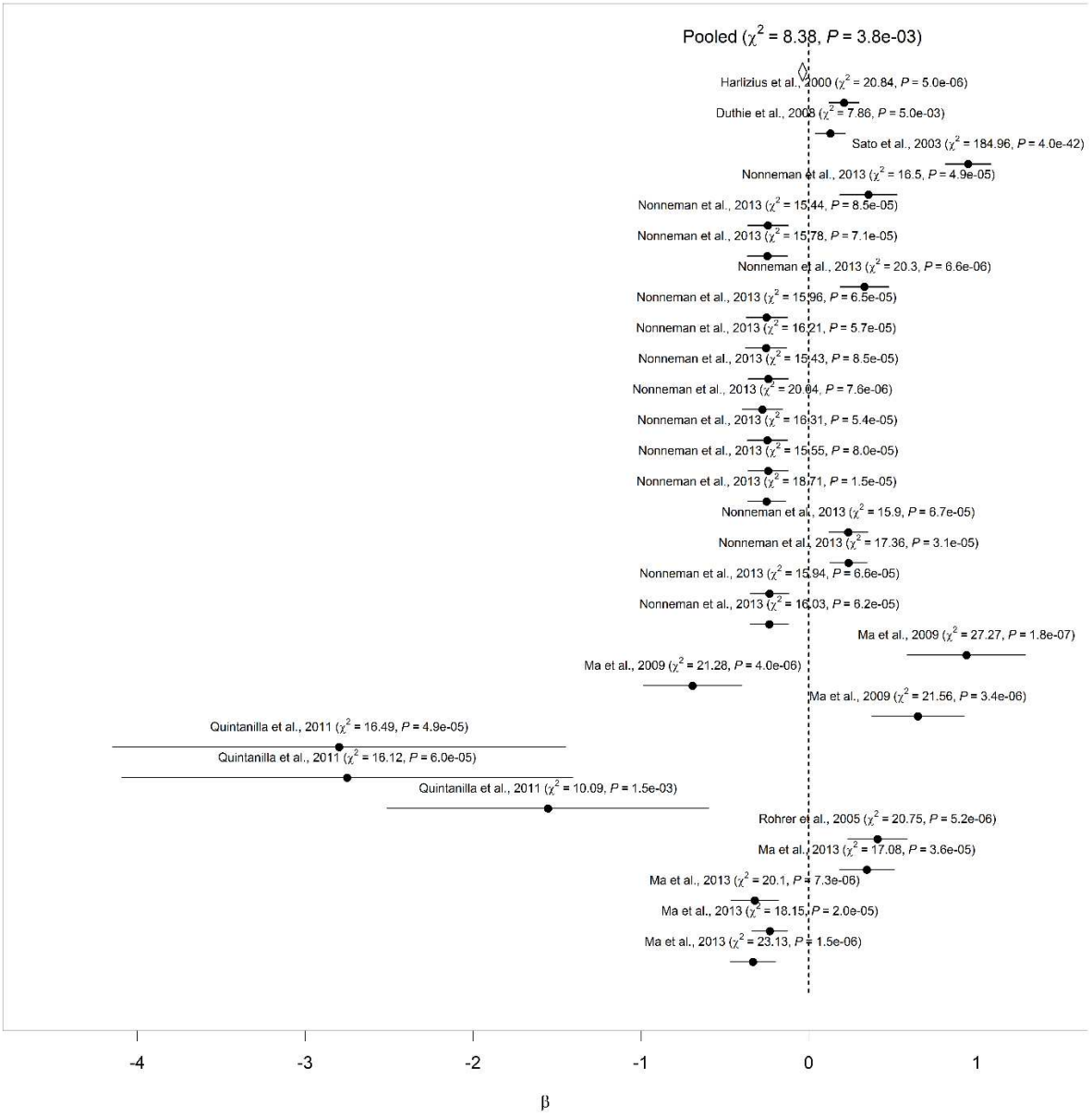

Figure 1 Forest plot from meta-analysis, with the p-value of each study. Through figure, we can see that all studies were significant (none of the horizontal lines cross the 'line of no effect')

Thus, we can compare the positions of significant marker from studies used in the meta-analysis and the positions of the significant SNPs found in the present study. In the Table 1, the position of the significant SNPs found in this study and the position of QTLs from meta-analysis is described. It is possible to note that some significant SNPs in this work are closer to QTLs from meta-analysis studies. Therefore, the meta-analysis confirm the results from genome association studies.

Table 1 Significant SNP markers (with respective genome positions, in bp, and NCBI related genes) from genome association studies of the present study and the significant SNPs and QTLs (with position) for IMF found and studies from meta-analysis

Genome Association Studies				**Meta-analysis**	
SNP	**Chr[1]**	**Position**	**Gene**	**QTL Studies**	**Position**
ALGA0025813	4	78563059	CLVS1	Nonneman et al., 2013	71524893
ALGA0025370	4	67382727	JPH1	Nonneman et al., 2013	68200145
ALGA0026242	4	89983155	DPT	Nonneman et al., 2013	86747423
ALGA0026769	4	100620769	KIRREL	Nonneman et al., 2013	105800661
ALGA0045997	7	134459770	ICK	Quintanilla et al., 2011	133000000*
				Quintanilla et al., 2011	133000000*
ALGA0041266	7	50028824	RHAG	Nonneman et al., 2013	51608179
ALGA0044298	7	110429708	SEL1L	Sato et al., 2003	113300000*
ALGA0048135	8	75754513	SDAD1	Nonneman et al., 2013	76669178
ALGA0047813	8	44927836	TLL1	Duthie et al., 2008	48700000*

[1]chromosome
* original unit centiMorgan (cM) convert to base pairs (1cM= 1 mega base)

Association Weight Matrix (AWM)

The SNPs from genome association studies were selected to build the AWM according to the criteria above (Materials and Methods). Then, the AWM had a total of 85 rows, being 48 genes and 37 SNPs (which has no close gene) and 12 columns corresponding to the traits.

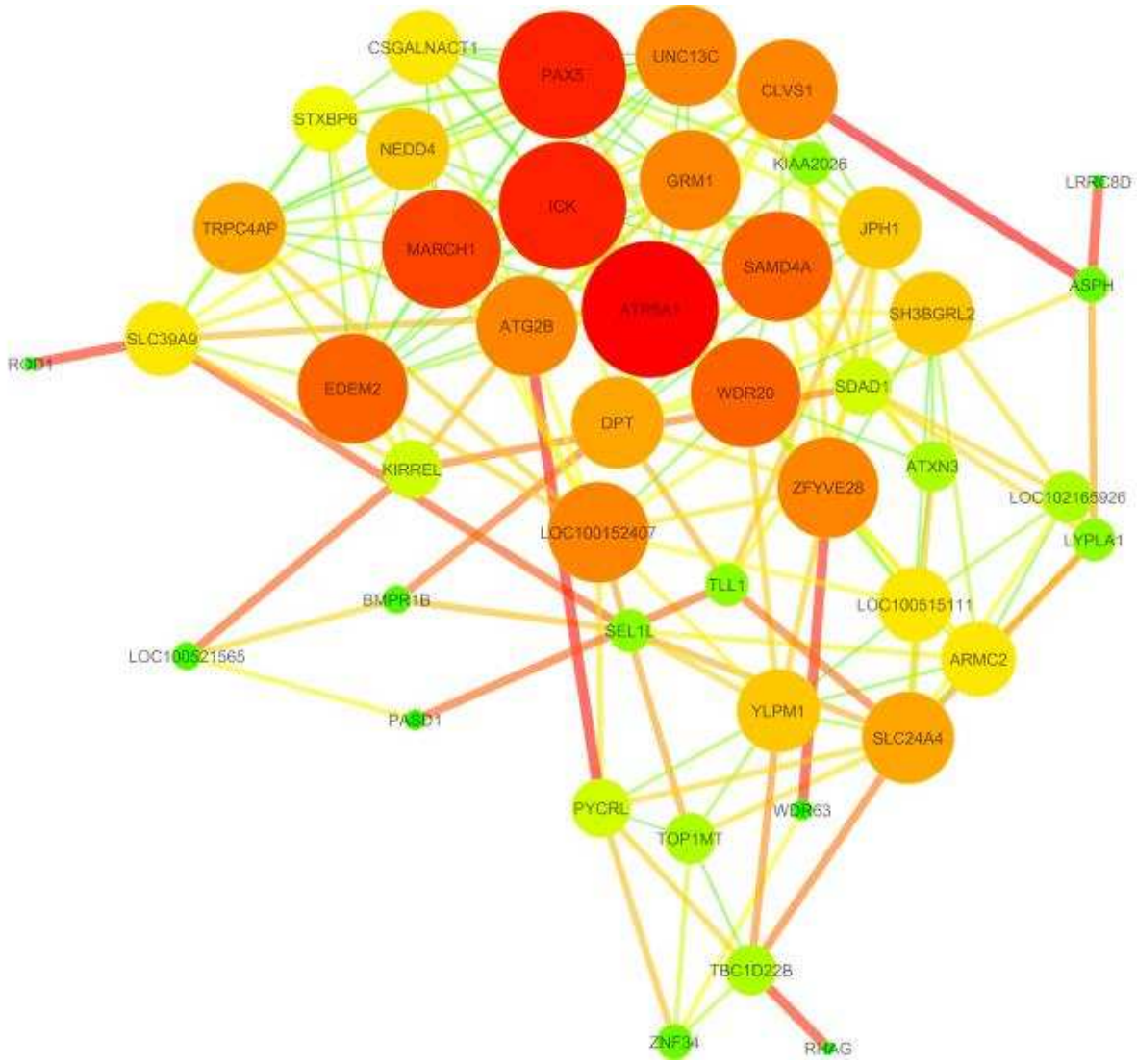

Figure 2 Correlated gene network predicted with association weight matrices (AWM) for meat quality and carcass traits.

The AWM was used to calculate pairwise correlations for the purpose of predict interactions between genes and hence build a gene network for meat quality and carcass traits. In the network, every gene was a node and every significant interaction was an edge connecting two nodes (Fig. 2). The genes with the most significant interactions are shown in the network with the stronger colors and the genes with more interactions (between them) are the largest nodes.

Transcription Factor

We identified 25 transcription factors (TF) strongly related (p-value<0.001) with genes in the network. The top three TF (Sox5, Nkx2-5 and T) were choosing for construction of a network with their pathways and gene ontology (Fig. 3).

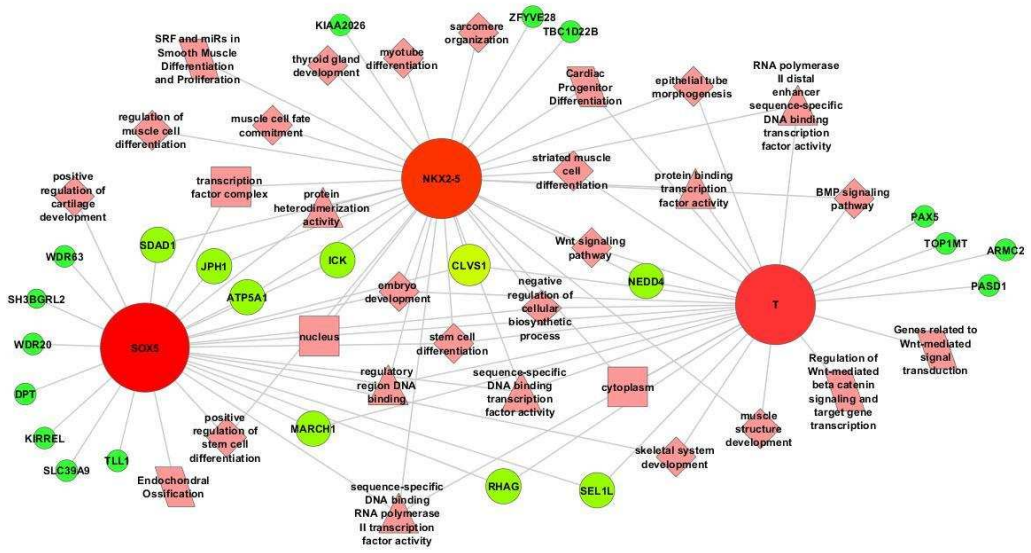

Figure 3 Transcription factors (in red) network, with related genes (green), pathways (parallelogram) and gene ontology: molecular functions (triangle), biological process (diamond) and cellular component (rectangle).

Discussion

We compared the position of our significant SNPs with the QTLs found in the meta-analysis studies, and we could note that some significant SNPs to IMF in this work are closer to QTLs from meta-analysis studies. For example, the SNP ALGA0025813 located at 78563059 bp in the chromosome 4 and the SNP found by Nonneman et al., 2013 at 71524893 bp in the same chromosome, the SNP ALGA0045997 located at 134459770 bp in the chromosome 7 and the SNP found by Quintanilla et al., 2011 at 133000000 bp (Table 1). Therefore, the meta-analysis confirm the results from genome studies association of the present study.

The predicted interactions calculated from standardized SNPs additive effects either reflect known biology or are hypotheses to be tested (Fortes *et al.*, 2010). AWM methodology has the advantage to handle more than one trait at the same time (considering the correlation between them). This fact increases the chances to detect genes that act directly or indirectly on a key trait, which would not be possible with the GWAS single-trait that focuses on the most significant SNPs for each trait at a time. Furthermore, we were able to confirm the relation between genes found in the network through the TF.

It is worth mentioning that the genes associated with the top TF (Fig.3) are those with more interactions (the bigger and darker nodes) in the network (Fig.2). From the top three TF chosen, SOX5 was related with 15 genes from the network. SOX5 encodes a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate. It is localized in QTL identified for meat quality (Ma et al., 2009, Nonneman et al., 2013). Besides, SOX5 has been associated with the regulation of myogenic progenitor cells and it was identified several potential SOX5-binding sites in the promoter of Pax7 (Rescan and Ralliere, 2010), which is involved in maintaining proliferation and preventing precocious differentiation of satellite cells from muscle skeletal (Zammit et al., 2006). Thus, it is possible that SOX5 is associated with the stem cells differentiation in the muscle skeletal tissue into muscle fibers instead of adipose tissue cells, indirectly influencing intramuscular fat. In the predicted network, SOX5 was associated with ATP5A1, this gene encodes a subunit of mitochondrial ATP synthase, known as a candidate gene for meat quality (Jennen et al., 2007) and muscular development (Große-Brinkhaus et al., 2010). Arakaki et al. (2007) working with

preadipocytes in mouse demonstrated that the cell-surface expression of ATP5A1 is markedly increased during adipocyte differentiation, suggesting that this gene has a role in lipid metabolism in adipocytes.

SOX5 was also associated with JPH1, a member of the junctophilin gene family, which encodes a protein component of junctional complexes. This gene contributes to the construction of the skeletal muscle triad by linking the t-tubule (transverse-tubule) and sarcoplasmic reticulum membranes. Another gene from network analysis associated with SOX5 was dermatopontin (DPT), which is an extracellular matrix protein with possible functions in cell-matrix interactions and matrix assembly. Dermatopontin is expressed in adipose tissue (Urs *et al.*, 2004, Keller *et al.*, 2008) and skeletal muscle, moreover it plays an important role in maintaining the balance of extracellular matrix metabolism, replacement and reconstruction of collagen (Zhang *et al.*, 2011). Again, we can suggest a relationship between SOX5 and the stem cells, since collagen is part of the connective tissue that is originated from the same mesenchymal stem cells that originates from skeletal muscle and adipose tissue.

NKX2-5 is known to regulate the expression of CARP (Zou et al., 1997, Kojic et al., 2011, Ma et al., 2013). The cardiac ankyrin repeat protein (CARP) is a MARP (muscle ankyrin repeat protein family) family member expressed in heart and skeletal muscle and it has different functions such as ion transport, regulation of cell cycle, initiation of transcription, cytoskeleton formation, signal transduction (Kojic et al., 2011). Thus, it is clear the importance of this protein in cells like the muscle cell with intensive metabolism. Therefore, CARP exhibits multiple functions, such as maintaining sarcomere structure, detecting and regulating intracellular and extracellular signals during muscle development (Ma et al.,

2013). Besides, CARP was found in fetal skeletal muscle, but barely detectable in adult skeletal muscle. It is a molecular marker in muscle from patients suffering from Duchenne muscular dystrophy (DMD) as well as other dystrophies (Witt et al., 2004). Studies have shown intramuscular fat increases in patients with DMD (Wren et al., 2008, LEROY-WILLIG et al., 1997). This is because in DMD patients, fat and collagen depositions are observed in skeletal muscle as dystrophic changes that progress with age, the fat accumulation and fibrosis originate from common mesenchymal progenitors (Uezumi et al., 2011).

Moreover, in a study with pigs, CARP was considered candidate genetic marker for meat quality (Ponsuksili *et al.*, 2009). Riazi et al. (2005) demonstrated that the expression of human NKX2-5 in C2C12 myoblasts inhibited myocyte differentiation and myotube formation and that increased expression of NKX2-5 in myotubes, normally expressing low levels of NKX2-5, leads to degradation of the sarcomere and division of myotubes into smaller myotubes.

The expression of NKX2-5 in terminally differentiated C2C12 myotubes resulted in a change in morphology and breakdown into smaller myotubes. In our network NKX2-5 was associated with nine genes, among them are ATP5A1 and JPH1 mentioned earlier associated with SOX5. This TF was also associated with ZFYVE28 (Zinc Finger, FYVE Domain Containing 28), a negative regulator of epidermal growth factor receptor (EGFR) signaling. ZFYVE28 localizes to early endosomes, promotes degradative sorting of activated EGFR (Mosesson *et al.*, 2009), a key regulator of myoblast differentiation. EGFR activity is down-regulated during early human myoblast differentiation, and this event is required for normal differentiation to take place (Leroy *et al.*, 2013).

Another interaction identified in this work is NKX2-5 with ICK. ICK encodes an intestinal serine/threonine kinase harboring a dual phosphorylation site found in mitogen-activating protein (MAP) kinases. The protein localizes to the intestinal crypt region and it is thought to be important in intestinal epithelial cell proliferation and differentiation. This gene has been associated with mTOR activity. Recent studies have demonstrated that in intestinal epithelial cells ICK deficiency led to a significant decrease in the mTOR activity (Fu *et al.*, 2009, Wu *et al.*, 2012). Activation of mTOR signaling is necessary for the induction of skeletal muscle hypertrophy (Bodine *et al.*, 2001, Goodman *et al.*, 2010). It has been proposed that mTOR may also control skeletal muscle mass through the regulation of protein degradation via the process of autophagy (Goodman *et al.*, 2011). Although ICK has not been associated with mTOR in skeletal muscle, it may play a role in mTOR activity in this tissue.

T was associated with PAX5 that encodes a member of the paired box domain transcription factors which plays a crucial role in cellular proliferation, differentiation and migration and thus tissue development. Rajan et al. (2012) have proposed PAX5 as a candidate gene for myoblast differentiation. PAX5 was also related as a novel loci of obesity, in a GWAS for human adolescent obesity (Melka *et al.*, 2012). The exact role of this gene in regulating adiposity is not clear, but Melka et al. (2012) suggested that it may be involved in sympathetically modulation. In mouse PAX5 postnatal expression is detected in periaqueductal gray matter of the midbrain and area postrema of the medulla oblongata. These areas of the central nervous system are part of the general sympathetic outflow to the periphery, including the adipose tissue in which sympathoactivation stimulates lipolysis. CLVS1 (clavesin 1) is also associated with T, and is required

for normal morphology of late endosomes and/or lysosomes in neurons. It also binds phosphatidylinositol 3,5-bisphosphate (PtdIns(3,5)P2) (Katoh *et al.*, 2009). PtdIns 3,4-P*2* may has involvement in mammalian cell responses to insulin by bind Akt. Akt is believed to be downstream of PI3K and important in insulin metabolic effects such as glucose transport and protein synthesis in 3T3-L1 adipocytes (Lawlor & Alessi 2001; SUMMERS et al. 1999; Ueki 1998). Since fatty acids are also synthesized from glucose, it is easy to see the importance of insulin in adipogenesis considering that insulin regulates glucose uptake by the cell. Furthermore, insulin stimulate fatty acids synthesis in liver when there are an excess of carbohydrate and decrease lipolysis through inhibition of hormone-sensitive lipase (Carvalheira et al. 2002; Foretz et al. 1999). According to current models, Akt is activated upon translocation to the plasma membrane and binding to PtdIns 3,4-P*2* and PtdIns 3,4,5-P3, generated on PI3K activation. This binding alters Akt conformation and eases the accessibility and phosphorylation by two upstream kinases, PDK1 and an unidentified kinase referred to as PDK2. PDK1, which is also activated by PtdIns 3,4,5-P3 binding (Ikonomov *et al.*, 2002). Besides, PtdIns(3,5)P2 has been associated with X-linked myotubular myopathy dysfunction in human (a genetic muscle degenerative disease) (Pendaries *et al.*, 2003, Michell *et al.*, 2006).

T was associated with MARCH1 and NEDD4. MARCH1 is a member of the MARCH family of membrane-bound E3 ubiquitin ligases. MARCH proteins add ubiquitin to target lysines in substrate proteins, thereby signaling their vesicular transport between membrane compartments. MARCH1 downregulates the surface expression of major histocompatibility complex (MHC) class II molecules and other glycoproteins by directing them to the late

endosomal/lysosomal compartment. (Bartee *et al.*, 2004, Thibodeau *et al.*, 2008). Even though MARCH1 was not associated with any of the present traits, MARCH6 a gene of the same family was associated with intramuscular fat metabolism (Liu *et al.*, 2010) and cholesterol homeostasis (Jiang and Song, 2014). NEDD4 encodes a E3 ubiquitin-protein ligase which accepts ubiquitin from an E2 ubiquitin-conjugating enzyme in the form of a thioester and then directly transfers the ubiquitin to targeted substrates. Jing Li et al. (2015) associated NEDD4 with obesity and insulin resistance. They showed that NEDD4 knockout mice (without expression of NEDD4) are partially resistant to insulin and under hight-fat-diet showed less body weight gain, less fat mass and smaller adipocytes vs. wild-type (with expression of NEDD4).

Conclusions

A meta-analysis confirm the results found in genome association studies in this present study. The AWM gene co-association network analysis revealed new associations between meat quality and carcass traits with genes mapped on SNPs significant. It was also possible to identify key transcription factors, gene-gene interactions and pathways underpinning the regulation meat quality and carcass traits. This is helpful to understand better the molecular basis involved with these traits. However, further efforts should be made in order to study in more detail the new gene-gene interactions here identified, as well as, the key transcription factors and pathways involvement in these traits.

References

Arakaki N, Kita T, Shibata H and Higuti T 2007. Cell-surface H+-ATP synthase as a potential molecular target for anti-obesity drugs. FEBS Lett 581, 3405-3409.

Band GdO, Guimarães SEF, Lopes PS, Schierholt AS, Silva KM, Pires AV, Benevenuto Júnior AA and Gomide LAdM 2005a. Relationship between the Porcine Stress Syndrome gene and pork quality traits of F2 pigs resulting from divergent crosses. Genetics and Molecular Biology 28, 88-91.

Band GdO, Guimarães SEF, Lopes PS, Peixoto JdO, Faria DA, Pires AV, Figueiredo FdC, Nascimento CSd and Gomide LAdM 2005b. Relationship between the Porcine Stress Syndrome gene and carcass and performance traits in F2 pigs resulting from divergent crosses. Genetics and Molecular Biology 28, 92-96.

Bartee E, Mansouri M, Hovey Nerenberg BT, Gouveia K and Fruh K 2004. Downregulation of Major Histocompatibility Complex Class I by Human Ubiquitin Ligases Related to Viral Immune Evasion Proteins. Journal of Virology 78, 1109-1120.

Bodine SC, Stitt TN, Gonzalez M, Kline WO, Stover GL, Bauerlein R, Zlotchenko E, Scrimgeour A, Lawrence JC, Glass DJ and Yancopoulos GD 2001. Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. Nature cell biology 3, 1014-1019.

Carvalheira JBC, Zecchin HG and Saad MJA 2002. Vias de Sinalização Da Insulina. Arq Bras Endocrinol Metabol 46 (4). ABE&M: 419–25.

Davoli R, and Braglia S 2008. Molecular approaches in pig breeding to improve meat quality. Brief Funct Genomic Proteomic 6(4), 313–332.

Duthie C, Simm G, Doeschl-Wilson A, Kalm E, Knap PW and Roehe, R 2008. Quantitative trait loci for chemical body composition traits in pigs and their positional associations with body tissues , growth and feed intake. Anim Genet 39, 130–140.

Fernandez X, Monin G, Talmant A, Mourot J and Lebret B 1999. Influence of Intramuscular Fat Content on the Quality of Pig Meat — 1. Composition of the Lipid Fraction and Sensory Characteristics of M. Longissimus Lumborum. Meat Sci 53 (1): 59–65.

Foretz M, Pacot C, Dugail I, Lemarchand P, Guichard C, Liepvre X, Berthelier-Lubrano C, Spiegelman B, Kim JB, Ferre P and Foufelle F 1999. ADD1/SREBP-

1c Is Required in the Activation of Hepatic Lipogenic Gene Expression by Glucose. Mol Cell Biol 19 (5): 3760–68.

Fortes MRS, Reverter A, Zhanga Y, Collisa E, Nagaraj SH, Jonssona NN, Prayaga KC, Barrisa W and Hawkena RJ 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. PNAS 107, 13642–13647.

Fu Z, Kim J, Vidrich A, Sturgill TW and Cohn SM 2009. Intestinal cell kinase, a MAP kinase-related kinase, regulates proliferation and G1 cell cycle progression of intestinal epithelial cells. American journal of physiology. Gastrointestinal and liver physiology 297, G632-G640.

Gao, SZ and Zhao SM 2009. Physiology, Affecting Factors and Strategies for Control of Pig Meat Intramuscular Fat. Recent Pat Food Nutr Agric 1(1): 59–74.

Goodman CA, Mayhew DL and Hornberger TA 2011. Recent progress toward understanding the molecular mechanisms that regulate skeletal muscle mass. Cell Signal 23, 1896-1906.

Goodman CA, Miu MH, Frey JW, Mabrey DM, Lincoln HC, Ge Y, Chen J and Hornberger TA 2010. A phosphatidylinositol 3-kinase/protein kinase B-independent activation of mammalian target of rapamycin signaling is sufficient to induce skeletal muscle hypertrophy. Mol Biol Cell 21, 3258-3268.

Große-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, Jungst H, Looft C, Schellander K and Tholen E 2010. Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc x Pietrain population. Genet Sel Evol 42, 39.

Habier D, Fernando RL and Dekkers JCM 2009. Genomic Selection Using Low-Density Marker Panels. Genetics 182, 343-353.

Harlizius B, Rattink AP, Koning DJ De, Faivre M, Joosten RG, Arendonk JAMV and Groenen MAM 2000. The X Chromosome harbors quantitative trait loci for backfat thickness and intramuscular fat content in pigs. Mamm Genome 11, 800–802.

Hidalgo AM, Lopes PS, Paixão DM, Silva FF, Bastiaansen JWM, Paiva SR, Faria DA and Guimarães SEF 2013. Fine mapping and single nucleotide polymorphism effects estimation on pig chromosomes 1, 4, 7, 8, 17 and X. Genetics and Molecular Biology 36, 511-519.

Hocquette JF, Gondret F, Baéza E, Médale F, Jurie C and Pethick DW 2010. Intramuscular Fat Content in Meat-Producing Animals: Development, Genetic and Nutritional Control, and Identification of Putative Markers. Animal 4 (2). Cambridge University Press: 303–19.

Ikonomov OC, Sbrissa D, Mlak K and Shisheva A 2002. Requirement for PIKfyve enzymatic activity in acute and long-term insulin cellular effects. Endocrinology 143, 4742-4754.

Jennen DGJ, Brings AD, Liu G, Jungst H, Tholen E, Jonas E, Tesfaye D, Schellander K and Phatsara C 2007. Genetic aspects concerning drip loss and water-holding capacity of porcine meat. Journal of Animal Breeding and Genetics 124, 2–11.

Jiang W and Song BL 2014. Ubiquitin ligases in cholesterol metabolism. Diabetes Metab J 38, 171-180.

Jing Li J, Ferry RJ, Diao S, Xue B, Bahouth SW and Liao F-F 2015. Nedd4 Haploinsufficient Mice Display Moderate Insulin Resistance, Enhanced Lipolysis, and Protection against High-fat Diet Induced Obesity. Endocrinology, en.2014-1909.

Jolliffe IT 1972. Discarding Variables in a Principal Component Analysis. I: Artificial Data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 21, 160-173.

Jolliffe IT 1973. Discarding Variables in a Principal Component Analysis. II: Real Data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 22, 21-31.

Katoh Y, Ritter B, Gaffry T, Blondeau F, Honing S and McPherson PS 2009. The clavesin family, neuron-specific lipid- and clathrin-binding Sec14 proteins regulating lysosomal morphology. J Biol Chem 284, 27646-27654.

Keller P, Petrie JT, De Rose P, Gerin I, Wright WS, Chiang SH, Nielsen AR, Fischer CP, Pedersen BK and MacDougald OA 2008. Fat-specific protein 27 regulates storage of triacylglycerol. J Biol Chem 283, 14355-14365.

Kojic S, Radojkovic D and Faulkner G 2011. Muscle ankyrin repeat proteins: their role in striated muscle function in health and disease. Crit Rev Clin Lab Sci 48, 269-294.

Lawlor MA and Alessi DR 2001. PKB/Akt: A Key Mediator of Cell Proliferation, Survival and Insulin Responses? J Cell Sci 114 (16): 2903–10.

Leroy MC, Perroud J, Darbellay B, Bernheim L and Konig S 2013. Epidermal growth factor receptor down-regulation triggers human myoblast differentiation. PLoS One 8, e71770.

Liu L, Yin J, Li W, Liu K, Peng Y, Tan P and Ma RZ 2010. Construction of a bacterial artificial chromosome library for the Rongchang pig breed and its use for the identification of genes involved in intramuscular fat deposition. Biochem Biophys Res Commun 391, 1280-1284.

Ma G, Wang H, Li Y, Cui L, Cui Y, Li Q, Li K and Zhao B 2013. Cloning, expression, and bioinformatics analysis of the sheep CARP gene. Mol Cell Biochem 378, 29-37.

Ma J, Yang J, Zhou L, Zhang Z, Ma H, Xie X, Zhang F and Xiong, X 2013. Genome-Wide Association Study of Meat Quality Traits in a White Duroc 6 Erhualian F2 Intercross and Chinese Sutai Pigs. PloS One 8(5).

Ma J, Ren J, Guo Y, Duan Y, Ding N, Zhou L, Li L, Yan X, Yang K, Huang L, Song Y, Xie J, Milan D and Huang L 2009. Genome-wide identification of quantitative trait loci for carcass composition and meat quality in a large-scale White Duroc x Chinese Erhualian resource population. Anim Genet 40, 637-647.

Melka MG, Bernard M, Mahboubi A, Abrahamowicz M, Paterson AD, Syme C, Lourdusamy A, Schumann G, Leonard GT, Perron M, Richer L, Veillette S, Gaudet D, Paus T and Pausova Z 2012. Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure. J Clin Endocrinol Metab 97, E145-150.

Michell RH, Heath VL, Lemmon MA and Dove SK 2006. Phosphatidylinositol 3,5-bisphosphate: metabolism and cellular functions. Trends Biochem Sci 31, 52-63.

Mosesson Y, Chetrit D, Schley L, Berghoff J, Ziv T, Carvalho S, Milanezi F, Admon A, Schmitt F, Ehrlich M and Yarden Y 2009. Monoubiquitinylation regulates endosomal localization of Lst2, a negative regulator of EGF receptor signaling. Dev Cell 16, 687-698.

Nonneman DJ, Shackelford SD, King DA, Wheeler TL, Wiedmann RT, Snelling WM and Rohrer GA 2013. Genome-wide association of meat quality traits and tenderness in swine. Journal of Animal Science 91, 4043-4050.

Pendaries C, Tronchère H, Plantavid M and Payrastre B 2003. Phosphoinositide signaling disorders in human diseases. FEBS Lett 546, 25-31.

Ponsuksili S, Murani E, Phatsara C, Schwerin M, Schellander K and Wimmers K 2009. Porcine muscle sensory attributes associate with major changes in gene networks involving CAPZB, ANKRD1, and CTBP2. Funct Integr Genomics 9, 455-471.

Puig-Oliveras A, Ballester M, Corominas J, Revilla M, Estellé J, Fernández AI, Ramayo-Caldas Y and Folch JM. 2014. A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. PloS One 9 (12): e114862.

Quintanilla R, Pena RN, Gallardo D, Cánovas A, Ramírez O, Díaz I, Noguera JL and Amills M. 2011. Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects 1. J Anim Sci 89, 2963–2971.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rajan S, Chu Pham Dang H, Djambazian H, Zuzan H, Fedyshyn Y, Ketela T, Moffat J, Hudson TJ and R. S 2012. Analysis of early C2C12 myogenesis identifies stably and differentially expressed transcriptional regulators whose knock-down inhibits myoblast differentiation. Physiological Genomics 44(2), 183-197.

Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R and Dehais P 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4, e6524.

Rescan PY and Ralliere C 2010. A Sox5 gene is expressed in the myogenic lineage during trout embryonic development. Int J Dev Biol 54, 913-918.

Reverter A and Chan EK 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics 24, 2491-2497.

Riazi AM, Lee H, Hsu C and Van Arsdell G 2005. CSX/Nkx2.5 modulates differentiation of skeletal myoblasts and promotes differentiation into neuronal cells in vitro. J Biol Chem 280, 10716-10720.

Rohrer GA, Thallman RM, Shackelford S, Wheeler T and Koohmaraie M 2005. A genome scan for loci affecting pork quality in a Duroc – Landrace F 2 population. Anim Genet 37, 17-27.

Sato S, Oyamada Y, Atsuji K, Nade T, Sato S, Kobayashi E, Mitsuhashi T, Nirasawa K, Komatsuda A, Saito Y, Terai S, Hayashi T and Sugimoto Y 2003. Quantitative trait loci analysis for growth and carcass traits in a Meishan × Duroc F2 resource population. J Anim Sci 81, 2938–2949.

Silva KM, Bastiaansen JW, Knol EF, Merks JW, Lopes PS, Guimaraes SE and van Arendonk JA 2011. Meta-analysis of results from quantitative trait loci mapping studies on pig chromosome 4. Anim Genet 42, 280-292.

Struchalin, M and Aulchenko, Y 2014. MetABEL: Meta-analysis of genome-wide SNP association results. R package version 0.2-0. http://CRAN.R-project.org/package=MetABEL

Summers SA, Yin VP, Whiteman EL, Garza LA, Cho H, Tuttle RL and Birnbaum MJ 1999. Signaling Pathways Mediating Insulin-Stimulated Glucose Transport. Annals of the New York Academy of Sciences 892 (1 THE METABOLIC): 169–86.

Thibodeau J, Bourgeois-Daigneault MC, Huppe G, Tremblay J, Aumont A, Houde M, Bartee E, Brunet A, Gauvreau ME, de Gassart A, Gatti E, Baril M, Cloutier M, Bontron S, Fruh K, Lamarre D and Steimle V 2008. Interleukin-10-induced MARCH1 mediates intracellular sequestration of MHC class II in monocytes. Eur J Immunol 38, 1225-1230.

Touzet H and Varre J-S 2007. Efficient and accurate P-value computation for Position Weight Matrices. Algorithms for Molecular Biology 2, 15.

Ueki K1998. Potential Role of Protein Kinase B in Insulin-Induced Glucose Transport, Glycogen Synthesis, and Protein Synthesis. J Biol Chem 273 (9): 5315–22.

Urs S, Smith C, Campbell B, Saxton AM, Taylor J, Zhang B, Snoddy J, Jones Voy B and Moustaid-Moussa N 2004. Gene Expression Profiling in Human Preadipocytes and Adipocytes by Microarray Analysis. Journal of Nutrition 134, 762-770.

Verardo LL, Silva FF, Varona L, Resende MDV, Bastiaansen JWM, Lopes PS and Guimarães SEF 2015. Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. Journal of Applied Genetics 56, 123-132.

Witt CC, Ono Y, Puschmann E, McNabb M, Wu Y, Gotthardt M, Witt SH, Haak M, Labeit D, Gregorio CC, Sorimachi H, Granzier H and Labeit S 2004. Induction and Myofibrillar Targeting of CARP, and Suppression of the Nkx2.5 Pathway in the MDM Mouse with Impaired Titin-based Signaling. Journal of Molecular Biology 336, 145-154.

Wu D, Chapman JR, Wang L, Harris TE, Shabanowitz J, Hunt DF and Z. F 2012. Intestinal cell kinase (ICK) promotes activation of mTOR complex 1 (mTORC1) through phosphorylation of Raptor Thr-908. Journal of Biological Chemistry 287(15), 12510-12519.

Zhang Y, Zan L and Wang H 2011. Screening candidate genes related to tenderness trait in Qinchuan cattle by genome array. Mol Biol Rep 38, 2007-2014.

Zou Y, Evans S, Chen J, Kuo HC, Harvey RP and Chien KR 1997. CARP, a cardiac ankyrin repeat protein, is downstream in the Nkx2-5 homeobox gene pathway. Development (Cambridge, England) 124, 793-804.

Stearns TM, Beever JE, Southey BR, Ellis M, McKeith FK and Rodriguez-Zas SL 2005. Evaluation of approaches to detect quantitative trait loci for growth, carcass, and meat quality on swine chromosomes 2, 6, 13, and 18. II. Multivariate and principal component analyses. J Anim Sci 83, 2471-2481.

Zammit PS, Relaix F, Nagata Y, Ruiz AP, Collins CA, Partridge TA and Beauchamp JR 2006. Pax7 and myogenic progression in skeletal muscle satellite cells. J Cell Sci 119, 1824-1832.

Data supplements

Table 2 Coefficients of traits with the 26 last principal components, which are less important to explain the total variation in 38 carcass traits

	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25
CW	0,0315	-0,0652	-0,0019	-0,0234	0,0239	0,0118	0,0160	0,0153	0,4651	-0,0720	-0,0486	0,1582	0,0195
RHCW	0,0025	-0,0398	-0,0119	0,0137	0,0515	0,0320	0,0755	0,0382	0,2567	-0,0720	-0,0574	0,1744	-0,0194
CY	0,1692	0,0862	-0,2340	0,1725	-0,1452	-0,1196	-0,0070	0,0367	0,2347	-0,0418	-0,0430	-0,2447	-0,0406
MBCC	0,1824	0,0686	-0,1654	-0,0202	-0,1862	-0,1197	0,0261	0,0649	0,1676	0,2080	0,0648	-0,1396	0,1467
MLC	0,1316	-0,0153	-0,1170	-0,0443	-0,2303	-0,2062	-0,0080	-0,0284	0,1619	0,1058	-0,0463	-0,1626	-0,0277
SBT	0,1743	0,3311	0,0970	-0,2011	-0,1717	0,2344	-0,5555	0,0872	0,1051	0,0568	0,0298	0,0552	-0,0044
LR	0,1620	-0,0902	0,0097	-0,0027	0,1816	-0,3138	**-0,3370**	0,2592	0,0996	0,2302	0,0186	-0,2523	0,2761
LL	0,1620	-0,0132	-0,0244	0,0088	0,2724	-0,1399	0,1427	-0,3177	0,0849	-0,1293	-0,1979	-0,1023	-0,0167
L	0,0883	-0,0837	-0,0652	0,1434	0,1619	-0,0371	0,1260	-0,0549	0,0813	0,0907	-0,0601	-0,2373	-0,1150
P2	-0,0037	-0,0990	0,0184	-0,1067	0,0435	-0,2188	0,0130	0,1101	0,0803	-0,0538	0,1575	0,2963	-0,1474
BCD	-0,1551	-0,0553	0,0178	-0,0618	-0,1523	0,1760	0,3290	0,3952	0,0741	-0,1480	0,2960	-0,2160	0,1982
LD	-0,0096	-0,1837	-0,2511	0,2928	-0,1809	-0,0099	0,0501	0,0762	0,0622	-0,2005	0,0229	0,0099	0,2364
LEA	-0,0232	-0,0576	-0,1659	0,0859	-0,0815	0,1660	-0,1365	0,1000	0,0264	0,3132	0,0752	**0,3039**	-0,3441
THW	0,0575	0,0611	0,0400	-0,1063	0,0778	0,2082	0,1606	-0,0740	0,0190	0,2192	-0,0470	0,1702	0,1620
HW	-0,0010	0,1308	0,1337	-0,1631	0,2057	0,3056	-0,0096	-0,1096	0,0174	0,1687	-0,1077	-0,1350	0,2234
TBSW	-0,3251	0,0329	-0,1365	0,1752	0,1172	-0,0757	-0,0666	0,1570	0,0165	-0,1242	-0,1141	0,0274	0,1412
BSW	-0,3664	0,0731	-0,1584	0,1086	0,0547	-0,0132	-0,0713	-0,1370	0,0128	0,0549	0,0403	-0,1021	-0,0625
TPSW	0,1354	-0,2056	-0,0426	0,0917	**0,3134**	0,0467	0,0851	0,1766	0,0084	-0,0688	0,0611	0,1945	-0,0311
PSW	-0,0670	-0,1921	0,2738	-0,1090	0,1391	-0,1430	-0,2075	-0,1880	0,0040	-0,1033	**0,5938**	-0,2173	-0,1983
TLW	-0,0432	0,3576	0,0028	-0,2830	-0,0871	-0,0382	0,2602	0,0195	0,0032	-0,2634	0,0842	0,1327	-0,1348
LW	-0,1426	0,1802	0,1692	-0,1419	-0,0718	-0,1377	-0,0480	**-0,3643**	0,0022	-0,3390	-0,0847	-0,2601	-0,0220
BCW	-0,2593	-0,1604	0,2074	0,2218	-0,2048	-0,0740	-0,3363	-0,0681	0,0015	-0,1784	-0,3661	0,1061	-0,1387
RW	**0,3938**	-0,3324	-0,2031	-0,2089	-0,2518	0,1253	-0,0278	-0,2537	-0,0047	-0,1595	-0,0471	0,0635	0,0537
JW	0,0479	-0,0313	-0,2112	0,1295	0,2681	-0,0451	-0,1251	0,0945	-0,0067	-0,0082	0,3320	0,0401	-0,0286
SLW	0,1282	-0,0573	0,4931	0,0662	-0,0980	**-0,2833**	0,2432	0,2900	-0,0203	0,2899	-0,2236	-0,0323	-0,0963
AF	-0,2507	-0,1225	0,0833	0,1102	-0,3975	0,2483	0,1349	-0,2466	-0,0568	**0,3621**	0,1706	-0,1873	-0,0960
pH45	-0,2126	-0,1377	-0,0663	-0,0661	0,1344	0,2573	-0,0951	0,1428	-0,0633	-0,0036	-0,1635	-0,1663	0,1334
pH24	0,0629	**0,5473**	-0,2373	0,3079	-0,0309	-0,0874	0,0472	-0,0257	-0,0752	0,0848	0,0646	-0,0224	-0,1064
L	0,2000	0,0389	0,0514	0,1736	0,1787	0,2819	0,0416	0,2650	-0,1046	-0,1376	-0,0921	-0,2925	**-0,5197**
A	0,0469	0,0168	0,0672	0,1453	0,1269	0,0418	0,0130	-0,0299	-0,1150	-0,0139	0,0149	0,0181	-0,0130
B	-0,0850	0,1310	0,1291	0,1065	-0,0253	-0,1464	0,0260	-0,1415	-0,1217	0,0952	0,1105	0,1633	0,2019
IMF	-0,0981	-0,1533	-0,0343	0,1351	0,1454	0,0636	0,0021	0,0471	-0,1413	0,1265	-0,0158	-0,0251	-0,0250
DL	0,1511	0,0733	0,1708	0,2427	-0,0584	0,2784	-0,0841	-0,0259	-0,1455	-0,1776	0,0441	-0,0604	0,2324
CL	-0,1720	-0,0452	**-0,2388**	-0,2214	0,0969	-0,1414	-0,0740	0,0028	-0,1552	0,0392	-0,0745	-0,0181	-0,0996
SF	0,2032	0,0106	0,2212	**0,3615**	-0,0171	0,0445	-0,0750	-0,0807	-0,1563	-0,1727	0,1729	0,1203	0,1163
TL	-0,0463	0,0052	-0,0926	-0,0352	0,0398	0,0399	-0,0953	-0,0107	-0,2745	-0,0620	-0,0312	-0,0435	0,0465
C	-0,0624	0,1269	0,1352	0,1775	0,0356	-0,0854	0,0406	-0,1554	-0,2977	0,0514	0,0918	0,1870	0,1563
H	-0,0651	-0,0021	-0,0560	-0,1386	-0,1250	-0,0843	-0,0040	-0,0022	**-0,4999**	0,0362	0,0040	0,0038	0,0718

Table 1 Coefficients of traits with the 26 last principal components, which are less important to explain the total variation in 38 carcass traits. Continuation

	PC26	PC27	PC28	PC29	PC30	PC31	PC32	PC33	PC34	PC35	PC36	PC37	PC38
CW	0,1210	0,0217	-0,0826	-0,0578	-0,0036	-0,0225	-0,0782	0,1456	**-0,7845**	-0,0291	0,3940	0,0006	0,0003
RHCW	0,1030	-0,0278	-0,0734	-0,0196	0,0114	-0,0299	-0,0817	0,0724	-0,1733	0,1949	**-0,8506**	-0,0272	-0,0001
CY	-0,0668	-0,0942	0,0419	0,0492	-0,0275	-0,0553	0,0153	0,0087	0,0203	-0,0062	0,0034	-0,0006	0,0000
MBCC	-0,1580	-0,1418	-0,0653	0,0590	0,4285	0,2491	**0,4293**	0,1118	-0,0124	0,0031	-0,0140	-0,0053	0,0001
MLC	-0,0697	0,1107	0,2503	0,0071	**-0,4401**	-0,2601	-0,3869	-0,0756	0,0739	0,0136	0,0214	0,0047	-0,0001
SBT	0,0044	0,0391	-0,0450	-0,0325	-0,0185	-0,0280	-0,0248	0,0564	0,0145	0,0194	-0,0280	0,0103	0,0001
LR	0,2267	0,0201	-0,1308	-0,0581	-0,0157	-0,0557	-0,0556	-0,0315	0,0149	0,0261	-0,0140	-0,0113	-0,0001
LL	-0,2437	0,1282	-0,2459	**0,5189**	-0,2061	-0,0830	0,2374	0,0093	-0,0106	-0,0026	0,0021	-0,0024	0,0001
L	-0,1485	0,0736	0,1745	-0,4635	0,3881	0,1170	-0,2920	-0,0949	-0,0115	-0,0098	-0,0095	-0,0052	0,0001
P2	-0,2151	-0,3378	**0,5444**	-0,1214	-0,2247	0,0653	0,3194	-0,0695	-0,0291	-0,0281	0,0014	0,0038	-0,0003
BCD	-0,0543	**0,4013**	0,2221	0,2140	0,1851	-0,1180	-0,0468	0,0292	0,0176	0,0198	0,0113	0,0034	-0,0001
LD	0,2452	-0,2941	0,0635	0,0291	-0,0130	-0,0569	0,0190	-0,0209	0,0166	-0,0008	0,0010	0,0053	0,0001
LEA	-0,3367	0,3080	0,0169	0,1517	0,0867	0,0577	-0,0692	-0,0078	-0,0199	-0,0040	0,0010	0,0026	0,0000
THW	0,2151	-0,0355	0,1751	-0,0184	-0,0489	-0,0196	-0,0247	0,5491	0,3809	-0,0651	0,1432	0,0149	-0,0001
HW	0,1002	-0,0289	0,2948	0,1421	0,0067	0,0421	0,1352	**-0,6371**	-0,0897	0,0184	-0,0049	-0,0037	-0,0003
TBSW	-0,0710	0,0382	0,0026	0,1368	-0,1323	**0,5974**	-0,2542	-0,0362	0,0920	-0,0265	0,0772	-0,0075	0,0001
BSW	0,0144	-0,0021	0,0897	-0,0285	0,1401	-0,5656	0,1972	0,0967	-0,0333	-0,0016	0,0250	0,0098	-0,0002
TPSW	-0,0385	0,2412	-0,3479	-0,3774	-0,1140	-0,1230	0,2202	-0,3204	0,2396	-0,0535	0,1031	0,0083	-0,0001
PSW	-0,0498	-0,1322	-0,0393	0,1931	0,0667	0,0717	-0,1496	0,0623	0,0207	0,0163	-0,0118	-0,0008	0,0004
TLW	0,0467	-0,3158	-0,2901	-0,0102	0,1572	-0,0849	-0,2412	-0,2063	0,1856	-0,0484	0,1288	0,0038	0,0001
LW	-0,0472	0,3297	0,0427	-0,3665	-0,2007	0,1769	0,3080	0,1871	0,0140	0,0079	-0,0414	-0,0056	-0,0002
BCW	0,1192	0,0179	0,0952	0,1680	0,2664	-0,0843	0,0489	-0,0673	0,2340	-0,0413	0,0938	0,0166	-0,0002
RW	0,0441	0,0524	0,0766	0,0826	0,0483	0,0961	-0,0515	0,0458	0,1323	-0,0258	0,0450	0,0011	0,0002
JW	0,0227	0,0067	-0,1049	-0,0153	-0,0456	-0,0369	0,0532	-0,0852	0,1005	-0,0111	0,0553	0,0046	0,0000
SLW	-0,1359	-0,1209	-0,0564	0,0407	0,0620	-0,0641	-0,0663	0,0232	0,0359	0,0110	0,0430	0,0031	0,0000
AF	0,1836	-0,1288	-0,2351	-0,0564	-0,3227	0,1664	0,1455	-0,0173	0,0079	0,0000	0,0037	-0,0016	0,0003
pH45	**-0,4389**	-0,2938	-0,1535	-0,1067	-0,0520	-0,0989	-0,0784	0,0823	0,0084	-0,0298	-0,0030	-0,0019	0,0001
pH24	-0,0007	-0,0107	0,0554	0,0416	-0,0739	0,0388	-0,0075	0,0071	-0,0041	0,0005	0,0091	-0,0007	0,0000
L	0,2344	-0,0954	0,0821	0,0555	-0,0181	0,0736	0,0658	0,0504	-0,0136	-0,0499	0,0179	0,0332	-0,0002
A	0,0454	0,0507	0,0193	-0,0065	-0,0046	-0,0023	-0,0211	-0,0051	-0,0228	-0,1530	-0,0548	0,6897	-0,0006
B	-0,0117	0,0933	-0,0194	0,0106	0,0216	-0,0164	-0,0361	-0,0281	-0,0427	**-0,6789**	-0,1507	-0,1132	0,0000
IMF	-0,0002	-0,0664	-0,0232	-0,0037	-0,1001	0,0185	0,0011	0,0049	-0,0122	-0,0031	-0,0004	-0,0060	0,0002
DL	-0,3992	-0,1735	-0,0096	-0,0706	-0,0727	-0,0965	-0,0669	0,0795	-0,0104	-0,0190	-0,0021	0,0017	-0,3809
CL	0,1162	0,0030	-0,0239	0,0038	0,0677	0,0025	0,0138	-0,0314	0,0250	0,0058	0,0048	-0,0035	-0,5371
SF	0,0724	0,0395	0,0573	0,0118	0,0297	-0,0081	0,0282	-0,0267	0,0037	-0,0144	0,0140	0,0016	-0,0001
TL	-0,1191	-0,0855	-0,0213	-0,0331	0,0116	-0,0475	-0,0237	0,0176	0,0128	-0,0059	0,0020	-0,0006	**0,7527**
C	-0,0272	0,0636	-0,0063	-0,0212	0,0189	0,0001	-0,0142	-0,0175	0,0582	0,6748	0,1445	0,0159	0,0003
H	-0,0576	-0,0314	-0,0359	0,0027	0,0061	0,0200	0,0074	-0,0073	-0,0089	0,0382	-0,0155	**0,7129**	-0,0006

Table 3 Number of SNPs significant for each trait

Traits	SNPs Significant
redness (A)	30
chroma (c)	15
intramuscular fat (IMF)	5
drip loss (DL)	12
cooking loss (CL)	21
carcass yield including feet and head (CY)	19
midline lower backfat thickness above the last lumbar vertebrae (L)	44
skinless and fatless boston shoulder weight (BSW)	21
total (bone-in) loin weight (TLW)	21
total jaw weight (JW)	19
total ham weight (THW)	26
bacon weight (BCW)	23

R script used in the construction of AWM

```
##Reading the p-value and effect files
AWM_P=read.table("AWM_P.txt")
AWM_A=read.table("norAWA.txt")
### Secondary SNP Selection: Select the SNPs that are associated with at least Ap phenotypes
Step1_select <- function(AWM_P){ AWM_P<=0.05 } ##### make a function to found which are smaller than 0.05
Step2 <- t(apply(AWM_P, 1, Step1_select))  ### apply the function
SignificantPhenotypes <- apply(Step2, 1, sum) ### calculates how many SNPs are significant for each trait
Step3 <- as.data.frame(SignificantPhenotypes)
snp=rownames(Step3)
Step3name=cbind(snp, Step3)
Step4Sorted = Step3name[order(SignificantPhenotypes,decreasing = TRUE),] ### SNPs rank

##Separate SNPs that will be placed in the AWM, without undergoing selection, which are those associated with the largest number of traits, for example, “n” SNPs
Step5_AWM=Step4Sorted[1:n,] ### these SNPs associated with more than two traits will the matrix
Step5=Step4Sorted[-(1:n),] ##### SNPs that will be selected yet
Step5SNP_IDs <- rownames(Step5)
Step5snp=as.data.frame(Step5[,1])
colnames(Step5snp) = c("snp")
Step5name=cbind(Step5SNP_IDs, Step5snp)
rownames(Step5name) <- Step5snp[,1]

###Classification by the position of SNPs
GenomeMap = read.table("Mappedall.txt", header=T, row.names=1)
snpmap=rownames(GenomeMap)
GenomeMap_=cbind(snpmap, GenomeMap)
data1=data.frame(Step5name)
```

```
colnames(data1)="snpmap"
SNPsofInterest=merge(GenomeMap_,data1, by='snpmap'); incomparables=NA
Step6SNPclose = subset(SNPsofInterest, Distance <= 2500)
Step6SNPfar = subset(SNPsofInterest, Distance> 1500000)
rownames(Step6SNPclose)=Step6SNPclose$snpmap
rownames(Step6SNPfar)=Step6SNPfar$snpmap
Step6SNPclose_IDs= as.list(rownames(Step6SNPclose))
Step6SNPfar_IDs=as.list(rownames(Step6SNPfar))
###1gene=1SNP
Step6 <- as.data.frame(Step5)
Step6close <- as.data.frame(Step6[rownames(Step6SNPclose),])
Step6Genes= data.frame(Step6SNPclose$feature_name, Step6close[,2] )
rownames(Step6Genes)=Step6SNPclose$snpmap
colnames(Step6Genes) = c("Gene", "SignifPhenos")
Step6Sorted = Step6Genes[order(Step6Genes$SignifPhenos,decreasing = TRUE),]
Step6Unique = subset( Step6Sorted, !duplicated(Step6Sorted$Gene))
Unique_SNP_IDs = as.list(rownames(Step6Unique))
###Preparing the files to calculate Ap (μ), the average number of traits associated to SNP that were significant (P < 0.05) for the key phenotype
Step6SNPfarid<-as.data.frame(rownames(Step6SNPfar))
colnames(Step6SNPfarid)<-c("snpmap")
Unique_SNP_ID = as.data.frame(rownames(Step6Unique))
colnames(Unique_SNP_ID)<-c("snpmap")
Step6SNPs = rbind(Unique_SNP_ID, Step6SNPfarid)
### Calculating the Ap
Step7SNPs <- subset(AWM_P, pvalue_IMF <=0.05) ### select only those SNPs that are associated with the key phenotype, e.g. IMF
snpmap<-rownames(Step7SNPs)
Step7SNPs<-cbind(snpmap,Step7SNPs)
Step7merg<-merge(Step6SNPs,Step7SNPs,by='snpmap')
###select the significant SNPs for a number of phenotypes>=Ap
Step7Ap <- subset(Step5, SignificantPhenotypes>=Ap)
Step7SNP_IDs <- as.data.frame(rownames(Step7Ap))
```

```
colnames(Step7SNP_IDs)<-c("snpmap")
###merge the two files
Step7SNPscl <- merge(Unique_SNP_ID, Step7SNP_IDs,by='snpmap')
Step7SNPsfar <- merge(Step6SNPfarid, Step7SNP_IDs,by='snpmap')
###Prepare to PCIT
rownames(Step7SNPscl)<-Step7SNPscl[,1]
Step8SNP_IDcl<- AWM_A[rownames(Step7SNPscl),]
rownames(Step7SNPsfar)<-Step7SNPsfar[,1]
Step8SNP_IDfar<- AWM_A[rownames(Step7SNPsfar),]
###replace the SNP name by the genes name
Step8gene<- Step6Unique[rownames(Step7SNPscl),]
Step8SNP_IDcl<-cbind(Step8gene[,1],Step8SNP_IDcl)
rownames(Step8SNP_IDcl)<-Step8SNP_IDcl[,1]
Step8SNP_IDcl<-Step8SNP_IDcl[,c(-1)]
Step8file<-rbind(Step8SNP_IDcl,Step8SNP_IDfar)
### Do the same with the n SNPs that was placed in the AWM, without undergoing selection
Step9SNP_IDs <- rownames(Step5_AWM)
Step9snp=as.data.frame(Step5_AWM[,1])
colnames(Step9snp) = c("snp")
Step9name=cbind(Step9SNP_IDs, Step9snp)
rownames(Step9name) <- Step9snp[,1]
data2=data.frame(Step9name)
colnames(data2)="snpmap"
SNPsofInterest=merge(GenomeMap_,data2, by='snpmap'); incomparables=NA
Step9SNPclose = subset(SNPsofInterest, Distance <= 2500)
Step9SNPout = subset(SNPsofInterest, Distance> 2500)
rownames(Step9SNPclose)=Step9SNPclose$snpmap
rownames(Step9SNPout)=Step9SNPout$snpmap
Step10 <- as.data.frame(Step5_AWM)
Step10close <- as.data.frame(Step10[rownames(Step9SNPclose),])
Step10Genes= data.frame(Step9SNPclose$feature_name, Step10close[,2] )
rownames(Step10Genes)=Step9SNPclose$snpmap
colnames(Step10Genes) = c("Gene", "SignifPhenos")
```

```
Step10Sorted = Step10Genes[order(Step10Genes$SignifPhenos,decreasing = TRUE),]
Step10Unique = subset( Step10Sorted, !duplicated(Step10Sorted$Gene))
Step11SNP_IDcl<- AWM_A[rownames(Step10Unique),]
Step11SNP_IDout<- AWM_A[rownames(Step9SNPout),]
Step11file<-rbind(Step11SNP_IDcl,Step11SNP_IDout)
###bind the 2 files to use in the PCIT
PCITfile<-rbind(Step11file,Step8file)
write.table(PCITfile,file="PCIT_imput.txt",row.name=TRUE, quote=F)
```

R script used in the PCIT package

```
library(PCIT)
data=read.table("PCIT_imput.txt",h=T,row.names=1)
data2=t(data)
m1=as.matrix(cor(data2))
# Perform PCIT on the correlation matrix
result <- pcit(m1)

# Get indices for the meaningful correlations
signif <- idx(result)
# Plot the distribution of meaningful
# correlations superimposed on all correlations
 plotCorCoeff(m1, list("PCIT Meaningful" =+ signif), col=c("red"))
# Get the indices for the non-meaningful correlations
 nonsignif <- idxInvert(nrow(m1), signif)
# Set non-meaningful correlations to zero
m1[nonsignif] <- 0
# Create an adjacency matrix from c e.g. by

el <- getEdgeList(m1)
# modify the edge list to include some useful attributes for cytoscape
el$sign[el$Weight<0] <- "-"
el$sign[el$Weight>0] <- "+"
el$Weight <- abs(el$Weight)
# write the edge list stuff to a file suitable for import into cytoscape
write.table(el, file="cyto_imput.txt", row.names=FALSE, col.names=TRUE, sep="\t", quote=FALSE)
```