UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Ciència Animal i dels Aliments

Facultat de Veterinària

CENTRE DE RECERCA EN AGRIGENÒMICA

Grup de Recerca de Genòmica Animal

GENETIC DISSECTION OF GROWTH AND MEAT QUALITY TRAITS IN PIGS

ANNA PUIG OLIVERAS

PhD Thesis in Animal Production

Bellaterra, September 2015

Supervisors

Dr. Josep Maria Folch Albareda

Dra. Maria Ballester Devis

El Dr. Josep Maria Folch Albareda, professor titular del Departament de Ciència Animal i dels Aliments de la Universitat Autònoma de Barcelona,

i

la Dra. Maria Ballester Devis, investigadora del Departament de Genètica i Millora Animal de l'Institut de Recerca i Tecnologia Agroalimentàries (IRTA).

CERTIFIQUEN:

Que **Anna Puig Oliveras** ha realitzat sota la seva direcció el treball de recerca "Genetic dissection of growth and meat quality traits in pigs" per obtenir el grau de Doctora per la Universitat Autònoma de Barcelona.

Que aquest treball s'ha dut a terme al Departament de Ciència Animal i dels Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona i a la unitat de Genètica Animal del Centre de Recerca en Agrigenòmica.

Bellaterra, a 15 de juliol de 2015

Dr. Josep Maria Folch Albareda Dra. Maria Ballester Devis

CONTENT

SUMMARY

Pork is the most consumed meat worldwide. The breeding programs have been working towards a sustainable pig production improving reproduction traits, growth and meat quality traits. The genetic basis determining these complex and economically important production traits remains elusive not being fully understood. In the present thesis we pursue to clarify the molecular mechanisms that determine growth and meat quality traits.

The functional role of *FABP4* and *FABP5* genes in determining the fatty acid composition in muscle and backfat tissues was evaluated. Animals inheriting the C allele for *FABP4:g.2634_2635insC* polymorphism showed higher intramuscular content of C16:0 and C16:1(n-7) fatty acids and, decreased content of C18:2(n-6) fatty acid and higher *FABP4* expression in backfat. Moreover, *FABP4:g.2634_2635insC* was located inside a *PPARG* binding site suggesting a role of this nuclear receptor in the regulation of *FABP4* gene expression. The *FABP5:g.3000T>G* SNP was the most significant marker for intramuscular percentages of C18:1(n-9), C18:2(n-6), and MUFA. However, this variant was not associated with *FABP5* gene expression, being *FABP5* gene expression in muscle regulated by other genomic regions located on SSC4, SSC6, SSC9 and SSC13.

Aiming to identify genes and pathways affecting the intramuscular fatty acid composition, the muscle transcriptome of two groups of pigs with extreme phenotypes for these traits was analyzed. A total of 131 genes mostly related to lipid metabolism pathways were identified as differentially expressed between groups. The functional analysis showed that animals with a higher content of PUFA presented low fatty acid and glucose uptake resulting in an inhibition of the lipogenesis.

Gene-by-gene approaches used until now are limited when analyzing complex traits usually implying several genes with small effects; moreover, they ignore functional interactions. For this reason, we decided to apply a gene network approach based on SNP-by-SNP co-association analysis to explore growth, conformation and fatness related traits. From the resulting network formed by 513 nodes and 639 edges, three transcription factors *PRDM16*, *ELF1* and *PPARG* were likely to be the major regulators of this network. Moreover, 54 genes identified within the network belonged to growth-related ontologies.

Finally, with the aim to evaluate functional candidate genes affecting intramuscular deposition and fatty acid composition traits from previous studies of our group, the mRNA expression of 45 genes was measured in 114 animals. The eGWAS identified 241 eSNPs distributed in 18 eQTLs. Three out of 18 eQTLs presented *cis*-acting variants and 16 eQTLs showed *trans* regulatory effects. The results highlighted putative key regulators and improved our knowledge in the functional regulatory mechanisms implicated in these complex traits.

RESUMEN

La carne de cerdo es la más consumida mundialmente. Los programas de mejora animal se han centrado en realizar una producción sostenible y eficiente mejorando tanto los caracteres de reproducción, como de crecimiento y calidad de la carne. Actualmente no se dispone de un conocimiento completo sobre la base genética que determina estos caracteres de producción y por este motivo el objetivo principal de esta tesis consistió en profundizar en los mecanismos moleculares que determinan los caracteres de crecimiento y calidad de la carne.

Los genes *FABP4* y *FABP5* fueron evaluados como genes candidatos para la composición de ácidos grasos en músculo y grasa dorsal. Los animales portadores del alelo C para el polimorfismo *FABP4: g.2634_2635insC* mostraron un mayor contenido intramuscular de ácidos grasos C16:0 y C16:1(n-7) y, un menor contenido de C18:2(n-6) y una mayor expresión del gen *FABP4* en grasa dorsal. Dicho polimorfismo se encuentra dentro de una diana de unión para el gen *PPARG* pudiendo este factor nuclear determinar las diferencias de expresión observadas. Respecto la expresión del gen *FABP5* en músculo, se detectaron regiones asociadas en los SSC4, SSC6, SSC9 y SSC13. Sin embargo, no se observó una clara asociación de genotipos del SNP *FABP5: g.3000T> G* con la composición de ácidos grasos.

Con el objetivo de explorar el transcriptoma del músculo para identificar genes y rutas metabólicas relacionadas con el contenido y la composición de ácidos grasos en músculo, se secuenció el ARN de dos grupos de animales que diferían en estos caracteres. Se identificó un total de 131 genes diferencialmente expresados entre grupos en su mayoría relacionados con las rutas del metabolismo lipídico. El análisis funcional mostró una menor captación de glucosa y una inhibición de la lipogénesis en los animales con un mayor contenido en PUFA.

Los enfoques utilizados hasta ahora para el análisis de los caracteres complejos implican generalmente varios genes con efectos pequeños y no consideran las interacciones funcionales. Por esta razón, se decidió aplicar un enfoque de redes génicas basada en el análisis de co-asociación entre SNPs para el estudio de caracteres relacionados con el crecimiento, la conformación y el engrasamiento. Se identificaron tres factores de transcripción, *PRDM16, ELF1* y *PPARG*, como principales reguladores de la red. Además, 54 de los genes identificados en la red pertenecían a ontologías relacionadas con el crecimiento.

Finalmente, con el objetivo de evaluar genes candidatos funcionales identificados en estudios previos de nuestro grupo afectando a caracteres de deposición y composición de ácidos grasos en músculo, se analizó la expresión de 45 genes en 114 animales. El eGWAS identificó 241 eSNPs distribuidos en 18 eQTLs. Tres de los 18 eQTLs presentaban variantes en *cis*, y en 16 de los eQTLs se identificaron zonas reguladoras en *trans*. Los resultados permitieron identificar potenciales reguladores y mejorar nuestro conocimiento sobre los mecanismos funcionales implicados en estos caracteres complejos.

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

ANNEXES

Paper I

Paper II

Paper III

LIST OF PUBLICATIONS

The present thesis is based on the work contained in the list of articles below:

- Paper I. **Puig-Oliveras A,** Ballester M, Castelló A, Gago M, Fernández AI, Folch JM. Analysis of *FABP4* and *FABP5* gene expression and polymorphisms affecting fatty acid composition. *(submitted in Genomics Selection Evolution)*

- Paper II. **Puig-Oliveras A,** Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, Casellas J, Folch JM, Ballester M (2014) Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. Plos One 2014; 9(7):e103668. doi: 10.1371/journal.pone.0099720.

- Paper III. **Puig-Oliveras A,** Ballester M, Corominas J, Revilla M, Estellé J, Fernández AI, Ramayo-Caldas Y, Folch JM (2014). A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. PLoS One. 2014; 9(12): e114862. doi: 10.1371/journal.pone.0114862.

- Paper IV. **Puig-Oliveras A,** *et al.* Expression-based GWAS identify variants, gene interactions and potential key regulators affecting the intramuscular content and fatty acid composition in porcine meat. *(preliminary manuscript)*

RELATED PUBLICATIONS BY THE AUTHOR

(Not included in the thesis)

- Corominas J, Marchesi JAP, **Puig-Oliveras A**, Revilla M, Estellé J, Alves E, Folch JM, Ballester M (2015). Epigenetic regulation of the *ELOVL6* gene is associated with a major QTL effect on fatty acid composition in pigs. Genet Sel Evol. 47(1): 20.

- Revilla M, Ramayo-Caldas Y, Castelló A, Corominas J, **Puig-Oliveras A**, Ibáñez-Escriche N, Muñoz M, Ballester M, Folch JM (2014). New insight into the SSC8 genetic determination of fatty acid composition in pigs. Genet Sel Evol. 46:28. doi: 10.1186/1297-9686-46-28.

- Corominas J, Ramayo-Caldas Y, **Puig-Oliveras A**, Estellé J, Castelló A, Alves E, Pena RN, Ballester M, Folch JM (2013). Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics 14:843. doi: 10.1186/1471-2164-14-843.

- Corominas J, Ramayo-Caldas Y, **Puig-Oliveras A**, Pérez-Montarelo D, Noguera JL, Folch JM, Ballester M (2013). Polymorphism in the *ELOVL6* gene is associated with a major QTL effect on fatty acid composition in pigs. PLoS One 8(1):e53687. doi: 10.1371/journal.pone.0053687.

ABBREVIATIONS

ACOX2 : acyl-CoA oxidase 2, branched chain

ACSM : acyl-CoA synthetase medium-chain family member

Akt : serine/threonine kinase

ALB : albumin

APOA : apolipoprotein A

AQP : aquaporin

AWM : association weight matrix

BC1_LD : IBMAP backcross between Iberian and Landrace breeds

CAST : calpastatin

ChIP-seq : chromatin-immunoprecipitation sequencing

CPT : carnitine palmitoyltransferase

CYP7A1 : cytochrome P450, family 7, subfamily A, polypeptide 1

CYP2C : cytochrome P450, family 2, subfamily C, polypeptide

C16:0 : palmitic acid

C16:1(n-7) : palmitoleic acid

C17:0 : heptadecanoic acid

C18:0 : stearic acid

C18:1(n-7) : octadecenoic acid

C18:1(n-9) : oleic acid

C18:2(n-6) : linoleic acid

C18:3(n-3) : α-linoleic acid

C20:0 : arachidic acid

C20:3(n-6) : eicosatrienoic acid

eGWAS : expression genome-wide association studies

eQTL : expression quantitative trait locus

ELF1 : E74-like factor 1

ELOVL6 : ELOVL fatty acid elongase 6

EP300 : E1A binding protein p300

FABP : fatty acid binding protein

FAS : fatty acid synthase

FAT : fatty acid translocase

FATP : fatty acid transport protein

FOS : FBJ murine osteosarcoma viral oncogene homolog

FOXP3 : forkhead box P3

GLUT : glucose transporter proteins

GWAS : genome-wide association studies

HIF1AN : hypoxia in ducible factor 1, alpha subunit inhibitor

HOXA2 : homeobox A2

HSL : hormone-sensitive lipase

IBMAP: Iberian × Landrace animal material

IGF2 : insulin-like growth factor 2

KIT : v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog

LMX1A : LIM homeobox transcription factor 1, alpha

MGLL : monoglyceride lipase

MUFA : monounsaturated fatty acids

NCOA : nuclear receptor coactivator

NGS : next generation sequencing

NR3C1 : nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)

PBX1 : pre-B-cell leukemia homeobox 1

PCIT : partial correlation coefficient with information theory

PEX2 : peroxisomal biogenesis factor 2

PIK3CG : phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma

PIK3R1 : phosphoinositide-3-kinase, regulatory subunit 1 (alpha)

PLA2G12A : phospholipase A2, group XIIA

PLIN : perilipin

PPAR : peroxisome proliferator activated receptor

PRDM16 : PR domain containing 16

PUFA : polyunsaturated fatty acids

QTL : quantitative trait locus

RGS4 : regulator of G-protein signaling 4

RNA-Seq : RNA sequencing

RT-qPCR : real-time quantitative polymerase chain reaction

RXR : retinoic X receptor

RYR1 : ryanodine receptor 1 (Skeletal)

SCD : stearoyl-CoA desaturase

SDHC : succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa

SFA : saturated fatty acids

SLC2A4 : solute carrier family 2 (facilitated glucose transporter), member 4

SNP : single nucleotide polymorphism

SSC : *Sus scrofa* chromosome

TBPL2 : TATA box binding protein like 2

USF1 : upstream transcription factor 1

UTR : untranslated region

WGS : whole genome sequencing

1. GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

1.1. Porcine meat production

Pork is the most widely consumed meat in the world representing the 36.4% of the meat produced [FAO 2013; http://www.fao.org/; accessed June 2015]. Thanks to the improvements in the pig breeding technologies, this is the fastest growing livestock sector in terms of sustainability and efficiency of meat production; reaching in 2013 approximately a billion and a half of produced animals, having a great economic impact (Figure 1.1). With 150 million pigs and about 22 million-ton carcass weight, the European Union is the second largest pork producer (22.3%) after Asia (57.3%) (Figure 1.2A), followed by North Central America (17.1%) [FAO 2013; http://faostat.fao.org/; accessed June 2015]. Moreover, the 14.7% of European pig production came from Spain [Eurostat 2015; http://ec.europa.eu/eurostat/; accessed June 2015]. Finally, the region in Spain with the highest number of pigs produced is Catalonia reaching the 7,457,000 animals in 2014 (Figure 1.2B) [Eurostat 2015; http://ec.europa.eu/eurostat/; accessed June 2015].

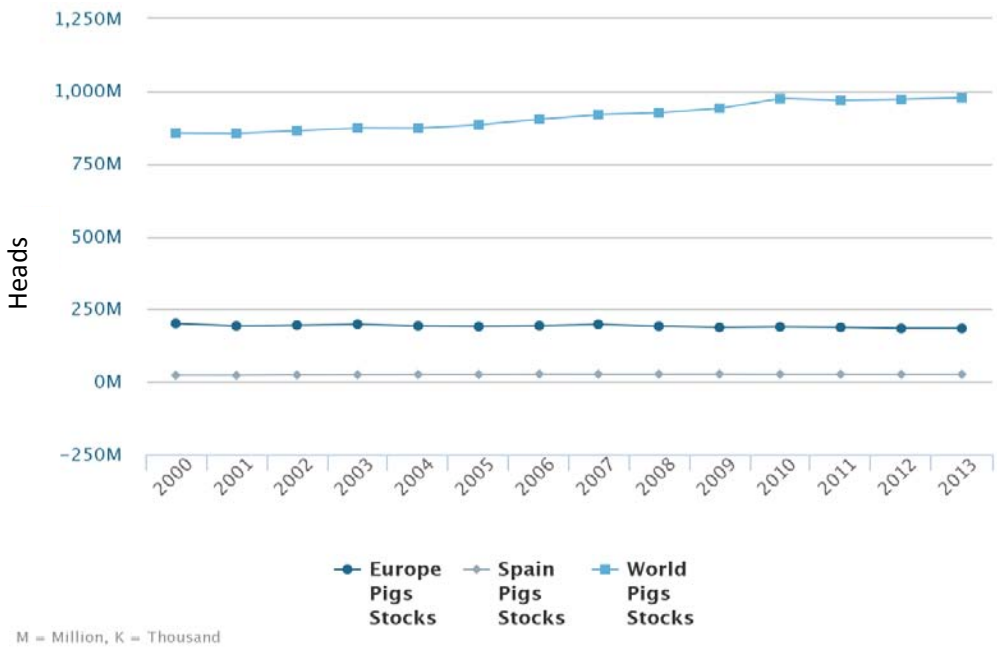


Figure 1.1. Evolution of pork meat production. Heads produced in the world, Europe and Spain from 2000 to 2013 [FAO 2013; http://faostat.fao.org/; accessed June 2015].

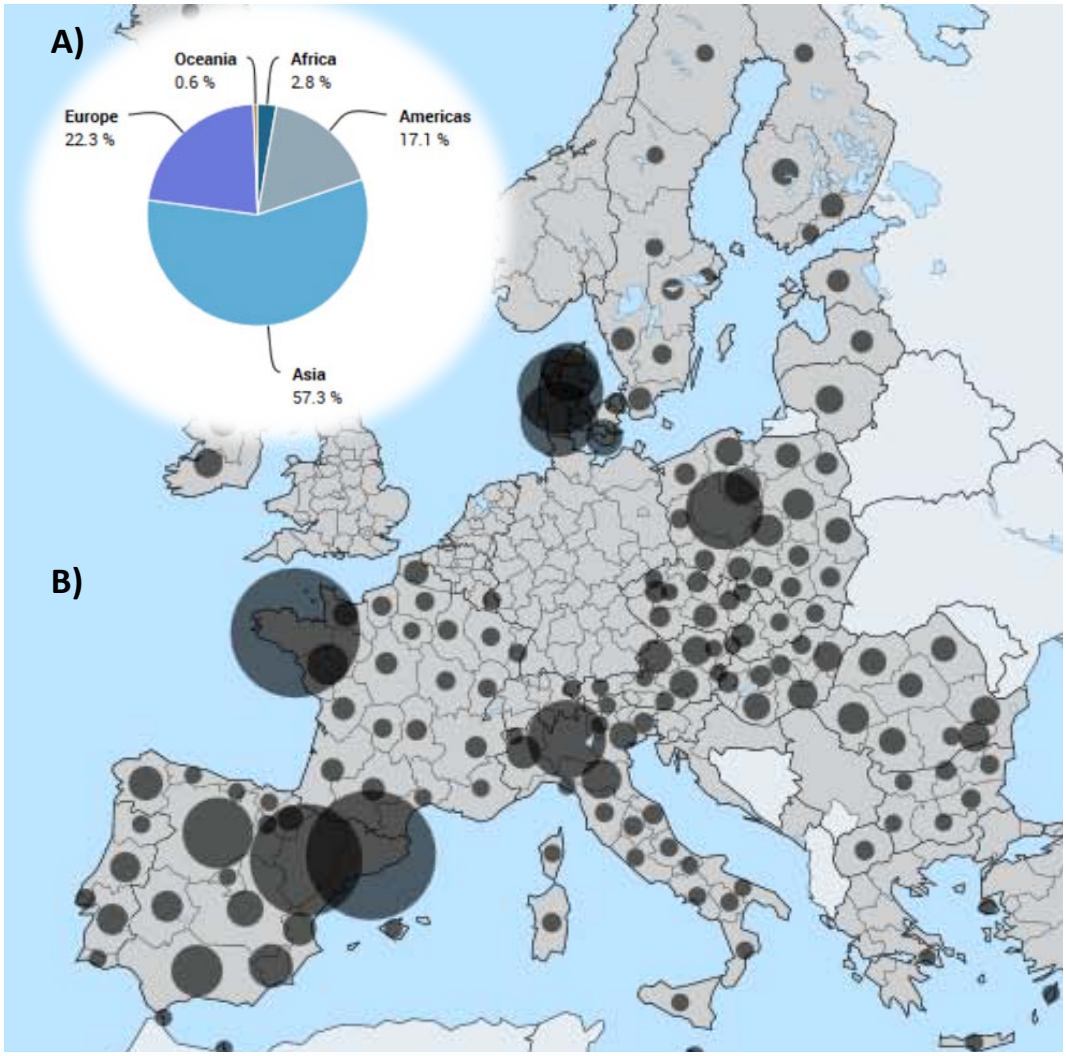


Figure 1.2. Importance of pig production worldwide and by regions. A) Pie chart representing the average number of pigs produced by continents in 2013 [FAO 2013; http://faostat.fao.org/; accessed June 2015]. B) Pigs produced by European regions in 2014 [Eurostat 2015; http://ec.europa.eu/eurostat/; accessed June 2015]. Circles size is according to the number of animals produced.

1.2. Relevant traits in the porcine industry

Pig breeders select measurable, attainable, realistic and timely goals to improve breeds according to the needs of producers, processors and consumers. The priority of these goals in the porcine selection has been in permanent revision changing along the time to satisfy demand (Dekkers *et al.,* 2011) (Figure 1.3). Over the last twenty years, swine breeding programs have focused on a few economic important

traits such as reproductive performance, growth efficiency and leanness, because there was an increasing demand for leaner pork at lowest price. As a consequence, a remarkable genetic progress was made, creating commercial animal breeds with a reduced backfat and high growth rate (e.g. Landrace). This contributed towards a massive utilization at a worldwide scale of these highly efficient breeds in an intensified and industrialized swine management, causing a declined importance of the native breeds (Kanis *et al.,* 2005; Pugliese & Sirtori 2012). However, the strong selection process lead to a dramatic reduction of intramuscular fat content in some breeds, negatively affecting meat quality, and therefore, leading to a deterioration of the taste and tenderness of the produced meat (Wood & Whittemore 2007).

For this reason, during the last years, consumers' requirements have changed, being more interested in healthier and tasty eating products (Wood *et al.,* 2004). With the aim to not only produce meat at a lower cost but also meet consumer's preference of high quality meat, concerning sensorial and nutritional aspects, pig breeding programs have recently included those traits in their goals (Figure 1.3).

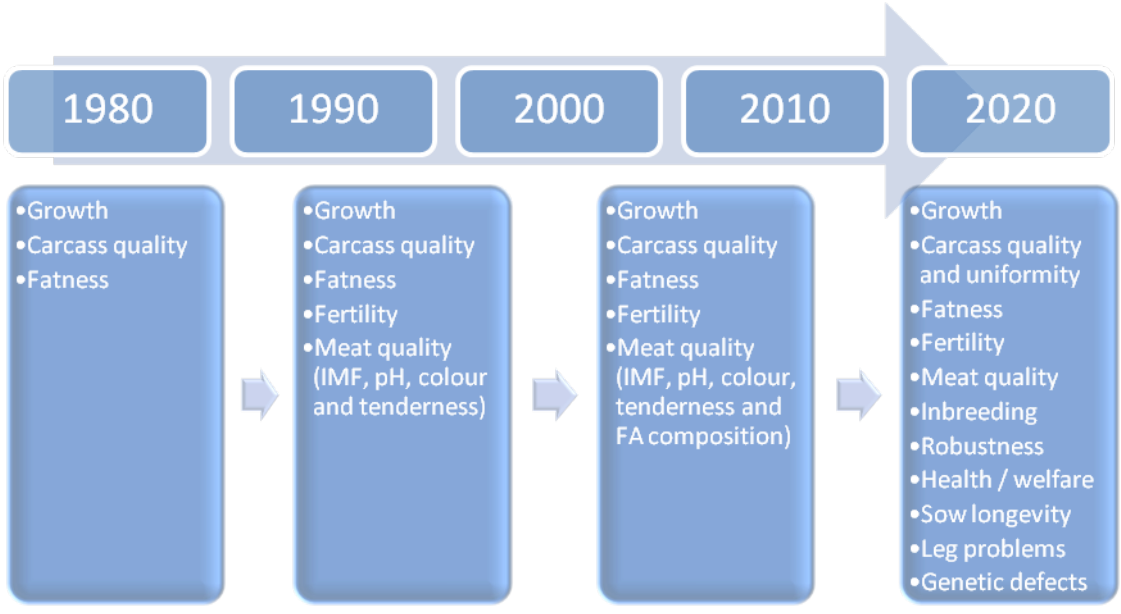


Figure 1.3. Main traits of interest for breeding in porcine meat production and their priority in selection along time (Toro & Silió, 1992; Ollivier 1998; Kanis *et al.,* 2005; Merks *et al.,* 2012).

General breeding interests consider growth (average daily gain, age at slaughter, feed efficiency), carcass quality (loin muscle area, primary cut weights), fatness (backfat thickness and abdominal fat), fertility (litter size, piglet weight and viability), disease resistance (immune capacity), behaviour (stress susceptibility, welfare), and meat quality traits. Meat quality comprises different aspects including technological (pH, water-holding capacity, cooking loss, firmness), sensorial (texture, flavour, juiciness, colour and marbling), nutrition (fat content and its lipid composition, digestibility) and safety (hygiene) factors (Figure 1.4).

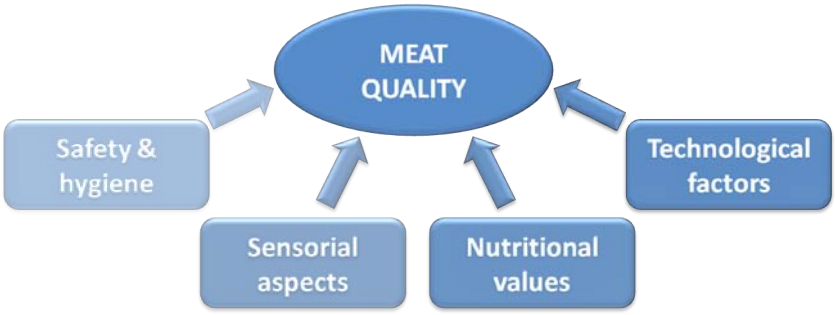


Figure 1.4. Schematic representation of the main factors that affect meat quality.

Hereby, there is a market requirement for high quality meat products, but at the same time, there is a strong request for low prices and safety products, which are critical factors for the food industry. Moreover, pig breeding has moved from the general genetic improvement of international lines to the specific lines for specific products (Pugliese & Sirtori 2012). This challenge can be addressed by traditional selection, management practices, and also by genomic selection (Miar *et al.,* 2015). Hence, pig breeding organisations need rapid genetic progress that will maximize value over costs without forgetting pig health and welfare. In the present thesis we focused on growth, carcass composition and conformation, and fat-composition related traits that are discussed below.

1.2.1. Growth, carcass conformation and composition

The carcass weight and yield, comprising the composition (proportions of muscle, fat and bone), are economically important traits in the pig meat industry affecting the value of different commercial cuts. The characteristics of the pig that are positive for profitability are high growth rate, low food conversion ratio and low fatness of the

carcass. Moreover, there is an increasing demand of the slaughterhouses for uniformity in the carcass cuts (Merks *et al.,* 2012). The major limitation for increasing slaughter weights faced two decades ago was the high carcass fat levels observed at heavier weights and the associated deterioration in feed efficiency (Cisneros *et al.,* 1996). The continued selection pressures have been successful at creating swine populations having heavier but leaner carcasses (Eggert *et al.,* 2007). Some carcass yield traits are relatively easy to measure in farms and slaughterhouses, such as body weight at different ages and at slaughter, the carcass length, the primary cuts weight such as ham, shoulder and belly, and the backfat thickness (Orcutt *et al.,* 1990; Whittaker *et al.,* 1992).

The differences in growth and carcass traits in pigs depend not only on breed, gender, maturity and environmental effects such as nutrition and management, but also have a strong genetic component (Davies & Kallweit 1979; Gu *et al.,* 1992; Eggert *et al.,* 2007). Genetic heritabilities for growth and carcass conformation traits are described in the literature to be moderate, ranging from 0.22 to 0.43 in different breeds (Johansson *et al.,* 1987; Hermesch *et al.,* 2000; Fernández *et al.,* 2007; Miar *et al.,* 2014). In the case of the proportions, the carcass lean percentage has an average heritability of 0.54 (Ducos 1994). This is not surprising since breeders have achieved through traditional selection a dramatic reduction of the amount of backfat present in carcasses.

1.2.2. Pork intramuscular fat content and fatty acid composition

The amount and composition of fat in muscle have gained special interest in the food industry for their close relation with meat tenderness and taste, being much appreciated for consumers, and giving an additional value to the final product (Wood *et al.,* 2004). Meat with high intramuscular fat is considered of good quality, giving flavour and oilness to the meat. In addition, intramuscular fatty acid composition affects the quality of meat in terms of palatability and nutritional values. Fat or adipose tissue is a connective tissue formed by adipose cells, which are constituted by 80% to 90% of lipids. Fatty acids, which are the predominant type of lipids, can be classified into three categories based on their saturation level: i) saturated fatty acids

(SFA) with no double bonds; ii) mono-unsaturated fatty acids (MUFA) with one double bond; and iii) polyunsaturated fatty acids (PUFA) with two or more double bonds. Fat properties depend on its fatty acid composition. The saturation of the fatty acids affects the fat firmness and oiliness changing the melting point. Fat with high PUFA content leads to faster oxidation rates producing a faster rancidity, whereas high MUFA contributes to a better taste and lower oxidation rate of meat. Furthermore, the ingestion of high SFA has been associated to obesity, cancers and cardiovascular diseases in humans. On the other hand, PUFA, mainly n-3, have been considered beneficial for human health, having a positive effect on cholesterol reduction and modulation of inflammatory processes. Therefore, there has been an increasing interest in the modification of intramuscular fatty acid composition for producing tastier and healthier meat (Wood & Enser 1997).

Both, fatty acid composition and deposition are complex characters determined by environmental factors such as diet and multiple genetic factors (Óvilo *et al.,* 2014, Ramayo-Caldas *et al.,* 2012a). Estimated heritability values for intramuscular fat range from 0.26 to 0.86, whereas for intramuscular fatty acid composition are low-to-moderate ranging from 0.15 to 0.55 (Casellas *et al.,* 2010; Ntawubizi *et al.,* 2010). The importance of the genetic determination of these traits has been evidenced in the comparison between pig breeds and the large amount of *quantitative trait loci* (QTLs) associated with these traits. However, the cost of measuring these traits and their low-moderate heritabilities make difficult to improve them by traditional selection (Davoli & Braglia 2007).

1.3. Fatty acid metabolism

Lipids are a major class of biological molecules and play many important and varied roles. They are crucial structural components of the membrane, are a major source of the energy stored in cells, and serve also as signaling molecules or lipokines (Cao *et al.,* 2008). The alterations that lipids undergo are produced mainly by two reactions, lipolysis or β-oxidation and lipogenesis or *de novo* fatty acid synthesis. The β-oxidation is a breakdown of the fatty acids that occurs in energy requirement states, whereas the *de novo* fatty acid synthesis is an energy storage mechanism.

Hence, fatty acids can be synthesized by *de novo* lipogenesis or provided by diet. In this regard, the fatty acid synthase (FAS) can synthetize palmitic acid (C16:0) from acetyl-CoA and malonyl-CoA which can be in turn the precursor of the long-chain saturated and unsaturated fatty acids of n-9 family (and minor fatty acids of n-7 and n-10 families); whereas, very long-chain fatty acid of n-3 and n-6 families are derived from α-linolenic acid (C18:3(n-3); an omega-3 fatty acid) and linoleic acid (C18:2(n-6); an omega-6 fatty acid). These two fatty acids are essential in human and other animals as they cannot be synthetized and must be provided by the diet (Guillou *et al.,* 2010) (Figure 1.5).

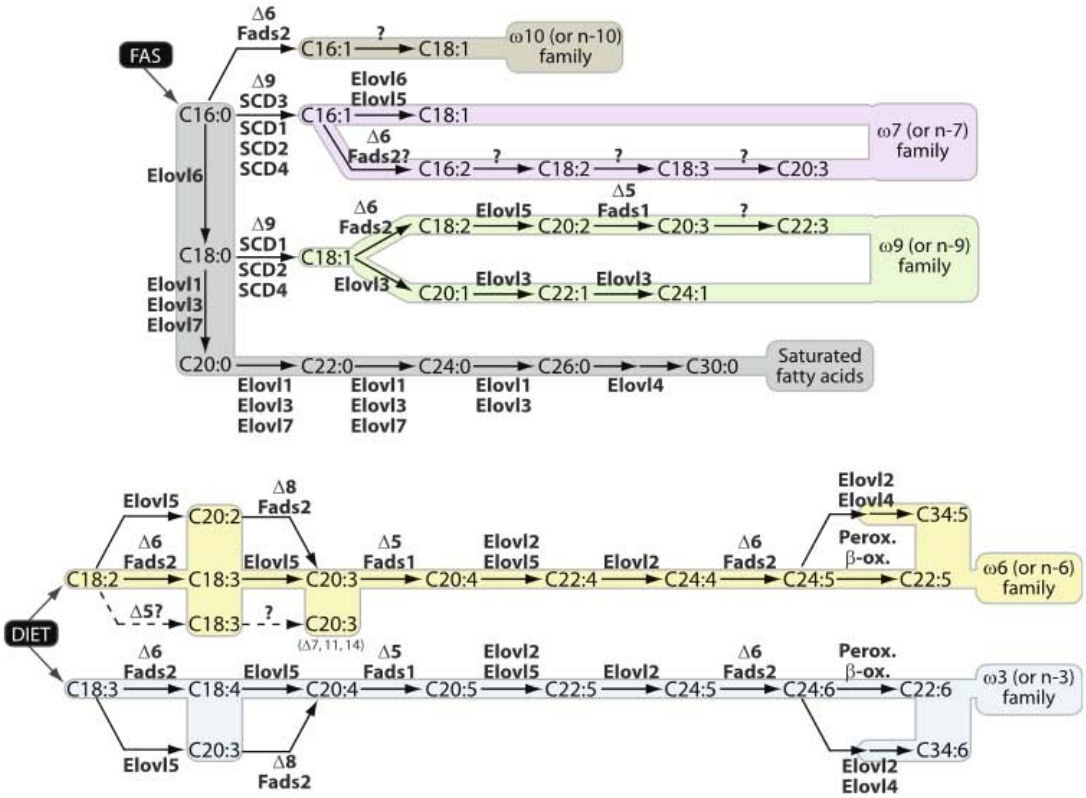


Figure 1.5. Fatty acid biosynthesis in mammals. The long-chain saturated and unsaturated fatty acids belonging to n-7, n-9 and n-10 families can be synthesized from C16:0 produced by FAS. On the other side, very long-chain fatty acids of n-3 and n-6 families are considered as essential because they can only be synthetized from precursors obtained from the diet (Guillou *et al.,* 2010).

The dietary fatty acid digestion begins on the stomach where lipids are partially digested by gastric lipases resulting in large fat globules with hydrophobic triacylglycerol cores surrounded by polar molecules, including phospholipids, cholesterol, fatty acids and ionized proteins. Afterwards, they move to the intestinal track where in the lumen they are hydrolyzed by the pancreatic lipase releasing monoacylglycerol, diacylglycerol and free fatty acids. Fatty acids and monoacylglycerol are absorbed by the enterocytes and are re-esterified to form triacylglycerol. Finally, chylomicrons formed from triacylglycerol together with cholesterol, phospholipids and proteins are transported via the lymphatic system to reach the target tissues where they are metabolized (Shi & Burn 2004). The main target tissues for lipid metabolism, the adipose tissue, liver, and skeletal muscle cooperate with each other to supply energy requirements. In adipose tissue, fatty acids are stored as triacylglycerol and are released into the circulation to meet demands of other tissues when necessary. In liver, they are re-esterificated into triacylglycerol and secreted as very-low-density lipoproteins. Conversely, in muscle they are utilized to obtain energy through oxidation. Several enzymes catalyze the different reactions of lipid metabolism (Figure 1.5 shows enzymes implicated in fatty acids biosynthesis). Furthermore, fatty acids in the circulation must pass the endothelium, the interstitial space, and the cell membrane for their utilization or storage. This transport of long chain fatty acids is mediated *via* specific transporter proteins including: the Albumin (ALB), which acts as a soluble fatty acid transporter in the blood plasma; the fatty acid transport proteins (FATPs) and the fatty acid translocase (FAT), which are plasma membrane proteins involved in the translocation of long-chain fatty acids from the intersticial space to the cytoplasm across the plasma membrane; and the fatty acid binding proteins (FABPs) involved in the fatty acid transport in the cell for their final oxidation to the mitochondria or storage in lipid droplets (Figure 1.6).

Moreover, there is a direct cross-talk between the glucose and fatty acid metabolism for energy homeostasis in the muscle (Kiens 2006). For instance, insulin acts in the carbohydrate metabolism facilitating the glucose diffusion into adipose and muscle cells via glucose transporter proteins (GLUT) and stimulates fatty acid synthesis and the storage of triglycerides by the esterification of glycerol phosphate. Finally, other enzymes and regulators influence the fatty acid metabolism such as the carnitine

palmitoyltransferases (CPTs) for fatty acid oxidation, the perilipins (PLINs) which protect fatty acids from their breakdown by the hormone-sensitive lipase (HSL), the peroxisome proliferator activated receptors (PPARs) which act as transcriptional regulators for genes involved in fatty acid metabolism among many others.

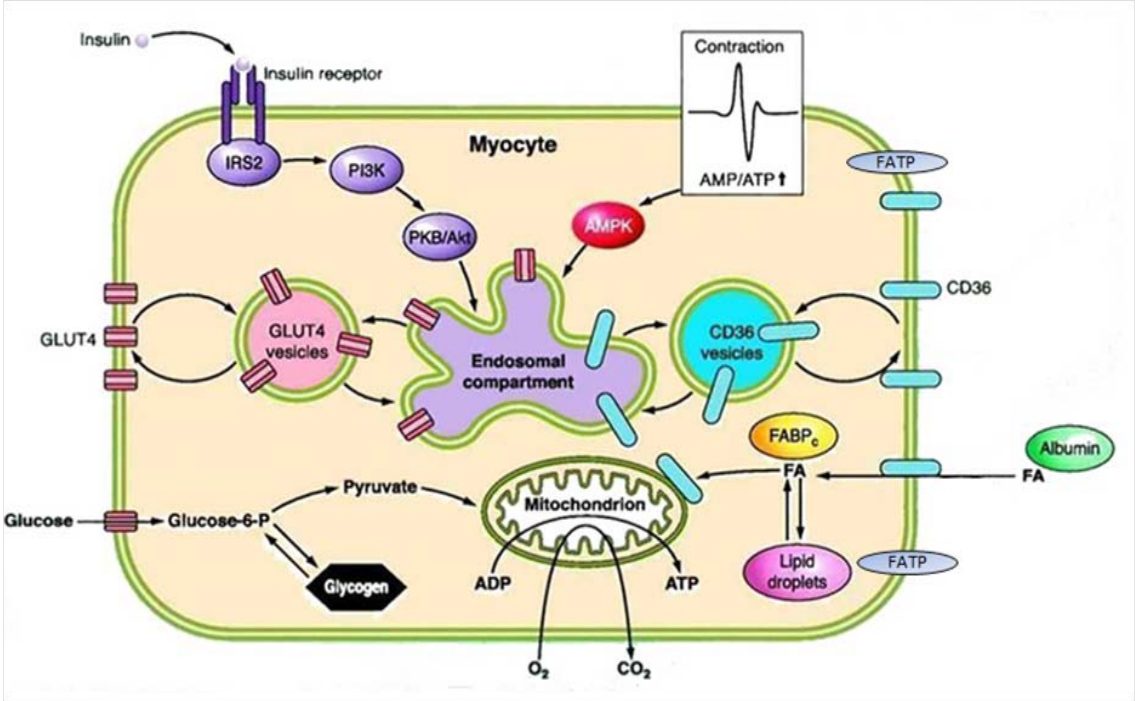


Figure 1.6. Fatty acid transport in the cell for utilization and/or storage in myocytes. Similarity between the regulation of cellular uptake of fatty acids and glucose. The uptake of both fatty acids and glucose by cardiac and skeletal muscle is increased after translocation of specific transporter proteins in response to stimulation with insulin or during increased contractile activity (adapted from Glatz *et al.,* 2010).

1.4. Genomic tools to study traits of interest in porcine production

The new genomic technologies are rapidly evolving and provide new opportunities to support breeding in a cost effective manner (Davoli & Braglia 2007) (Table 1.1). In swine species many tools have been developed by the scientific community to support the study of the production traits. In the nineties, the first map of genetic markers in the porcine genome was released (PiGMaP; Haley *et al.,* 1990). This served as a tool to identify QTLs through linkage analyses (Andersson *et al.,* 1994). Some years after, a big effort was made to sequence the pig genome (Groenen *et al.,* 2012) of about 2,596 Mb size from a single female Duroc animal (Archibald *et al.,* 2010), which has been submitted to continuous improvement. The most recent upgraded version of the pig genome sequence, named *Sscrofa10.2,* was released three years ago (Groenen *et al.,* 2012). The last upgraded annotation available for this sequence in the *Ensembl* database is the release 80 [Ensembl; http://www.ensembl.org, accessed May 2015], which comprises a total of 21,640 genes encoding 30,585 transcripts, 3,125 non-coding genes and 568 pseudogenes. Moreover, there is information about more than 52 millions of short variants, including single nucleotide polymorphisms (SNPs), insertions and deletions, and 85 structural variants. In addition, the development of high throughput SNP chips in porcine species such as the PorcineSNP60 BeadChip (*Illumina*) offered the possibility to perform high-density genotyping in a large number of animals, providing maps of higher resolution (more than 64,232 SNP markers) at a reduced cost. This resolution would be increased in a short time with the availability of new porcine SNP chip with more than 600,000 SNPs (Axiom® Porcine Genotyping Array, *Affimetrix*).

On the other hand, the development of the microarray technology allowed the expression analysis of many genes in a single reaction, quickly, and in an efficient manner. The first commercially available pig microarray (Porcine AROS v1.0, *Operon*; Gene-Chip® Porcine microarray, *Affymetrix*) was released in 2003 and consisted of a set of 10,665 oligo set (reviewed in Pena *et al.,* 2014). Some years later, these arrays were improved and, currently, they offer a more exhaustive coverage of the transcriptome and allowed to customize gene expression arrays. Different microarray studies have been performed in the porcine species to compare the muscle transcriptome of individuals with different intramuscular fat content and

composition (Liu *et al.,* 2009; Cánovas *et al.,* 2010; Damon *et al.,* 2012; Hamill *et al.,* 2013; Pena *et al.,* 2013; Sun *et al.,* 2013; Yu *et al.,* 2013) and during growth (D'Andrea *et al.,* 2011).

Recently, the development of high-throughput array platforms using nanoliter fluidics technology with customized oligonucleotide designs (i.e.: the Fluidigm Dynamic Arrays (*Fluidigm*) or the Taqman OpenArray plates (*Life technologies*)) allow genotyping or studying the expression of a low number of pre-selected DNA markers or genes, respectively, in large populations in a cost-effective manner.

Afterwards, the development of next generation sequencing (NGS) technologies to study whole genomes (whole genome sequencing or WGS) and transcriptome architectures (RNA sequencing or RNA-Seq) have represented a significant improvement in the analysis of complex traits (discussed in more detail in section 1.4.2). RNA-Seq enables to determine the transcript abundance with a larger dynamic range of expression levels compared with microarrays and it is not limited by the available genomic sequencing information during microarray production.

Table 1.1. Description of the main technologies that allow system genetics for quantitative traits and potential applications (adapted from Mackay *et al.,* 2009).

Layers or level of "omics"	Technology available	Potential applications
High-density genotyping	- Commercial genotyping arrays - Customized genotyping arrays	Identify SNPs in large number of individuals, identification of QTLs, structural variants and construction of networks based in co-association among SNPs
DNA sequencing	- Whole genome sequencing - Whole-exome sequencing - Sequence capture - Microbiome sequencing	Identification of variants (SNP, indels, structural, rare alleles), resequencing of interesting regions (QTLs, genes), analysis of microbiome and its interaction with the phenotypes
Transcriptional profiling	- Microarrays - RNA-Seq - High throughput qPCR microfluidic systems	Differential expression among groups, identification of variants, prediction of new transcripts and non-annotated genes, analysis of gene expression in larger populations, identification of eQTLs, co-expression analysis and networks
Epigenome	- Bisulfite sequencing - Chromatin immunoprecipitation sequencing - DNase I hypersensitive sites sequencing - Formaldehyde-assisted isolation of regulatory elements sequencing - Chromosome conformation capture	Analyze methylation and histone modification patterns, chromatin architecture and its interaction with proteins
Proteomics	- Tandem mass spectrophotometry	Detection of quantitative and qualitative variation on proteins
Metabolomics	-Gas chromatography -High-performance liquid chromatography-mass spectrophotometry	Detection of quantitative and qualitative variation in cellular metabolites
Phenomics	- Image or video analysis-based phenotyping	Phenotyping large number of samples required for system genetics analyses

However, the main RNA-Seq drawback when compared with microarrays is that the analysis relies on the current pig genome assembly. In swine, different RNA-Seq studies have been performed concerning growth and meat quality traits (Chen *et al.,* 2011; Pérez-Montarelo *et al.,* 2012; Ramayo-Caldas *et al.,* 2012b; Corominas *et al.,* 2013b; Jiang *et al.,* 2013; Puig-Oliveras *et al.,* 2014a; Xing *et al.,* 2014).

It is worth mentioning that also other new developed high-throughput technologies allow the analysis of the whole proteome or metabolome (Table 1.1), which is very convenient because of their effect on meat quality traits (Paredi *et al.,* 2013). For instance, in pigs, proteome studies have been used to identify protein modifications in muscle during *post mortem* ageing in relation with meat quality traits (Di Luca *et al.,* 2013). Last but not least, new technologic advances have made possible to phenotypically characterize a large amount of samples in a reduced time and in a more easy way (phenomics), for example by using the Fat-o-Meter for carcass grading system based on lean content with AutoFOM device in pigs (Mohrmann *et al.,* 2006). The large-scale genetic data obtained with these methodologies have required the development of new and more efficient computational technologies, which has boosted the bioinformatics area in the scientific community.

Finally, the integration of different layers of biological information, the so-called "omics" or systems genetics which takes into account the genomics, transcriptomics, proteomics, metabolomics, etc., is a major challenge to decipher the genetic basis of variation for quantitative traits (Civelek & Lusis 2014, Ohashi *et al.,* 2015). This combined approach may be useful to understand the flow of biological information that underlies complex traits. To achieve this objective, multi-dimensional high-density omics data have to be integrated using mathematical or statistical models, computational biology and bioinformatic tools (Kadarmideen 2014). Few pionnering studies in porcine species have applied systems biology approaches. For instance, Yang *et al.* (2011), with the aim to study the relationship between five endocrine tissues and 27 plasma metabolites, integrated the analysis of transcriptome and metabolome to identify quantitative trait transcripts. Similar approaches such as the integration of genomics and transcriptomics to identify heritable *expression quantitative trait loci* (eQTL; more discussed below) have been also performed in swine species (Cánovas *et al.,* 2012; Muñoz *et al.,* 2013).

1.4.1. QTLs, GWAS and candidate genes

The final goal of QTL and genome-wide association studies (GWAS) in domestic animals is to identify genes and variants associated with economically relevant traits. To achieve this objective, these techniques use a quantitative phenotype and screen for association among markers, such as microsatellites or SNPs distributed along the genome, and phenotypic records. The QTL approach relies on the existence of predictive markers near or linked to causal loci that tend to segregate together. It is useful for populations with related individuals where the tendency of this linkage is disrupted by recombination events, which probability increases with physical distance. The most predictive markers are therefore expected to reside in the proximity of the causal locus (Mackay *et al.,* 2009). The succes of QTL identification is affected by (1) the allele segregation in the cross and (2) the number of recombination events; requiring large data sets of related individuals with known pedigrees (Hill 2010). On the other hand, the GWAS allow assessing the association of each separate marker for the trait of interest in populations and crosses. Several thousand markers spaced throughout the genome, usually SNPs, are used to identify a marker allele in linkage or linkage desequilibrium with the causal variant. Therefore, if one allele is more frequent in animals showing a certain phenotype, it is defined as an associated SNP for this trait and considered as a marker. This screening serves as an information source for breeding value estimation and it is used to assess genomic selection in the industry. Genomic selection is predicted to offer anywhere between 20-50% of greater genetic progress by using pig breeding programs (Meuwissen *et al.,* 2001; Huisman & Charagu 2013).

Currently, there have been reported 13,030 pig QTLs in the *Pig QTLdb* represented by 663 different pig traits [PigQTLdb; http://www.animalgenome.org/cgi-bin/QTLdb/SS/index; accessed July 2015]. From the total QTLs reported, 1,880 have been detected for fatness traits, 646 are related to fat composition traits and 1,070 and 710 to growth and conformation traits, respectively (Figure 1.7).

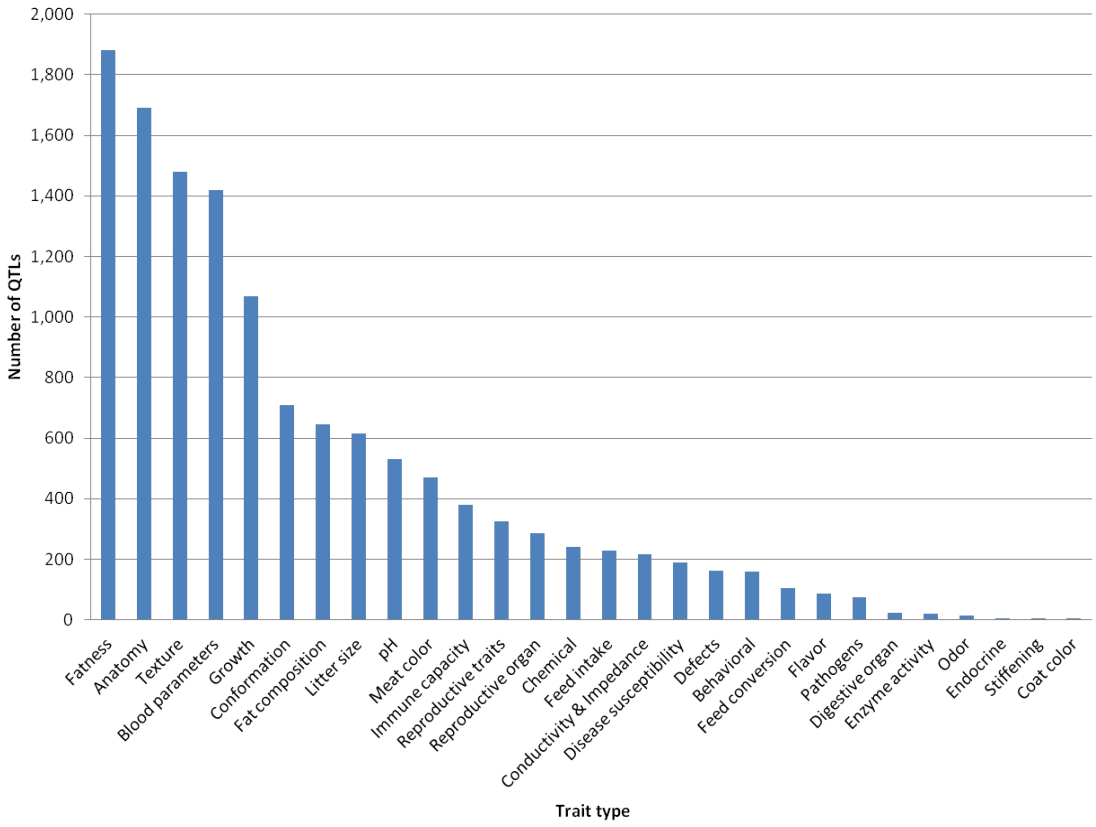


Figure 1.7. Distribution of porcine QTLs among the different trait types in the Pig QTLdb [PigQTLdb; http://www.animalgenome.org/cgi-bin/QTLdb/SS/index; accessed July 2015].

Noteworthy, the efforts in QTL and GWAS analyses have been useful to identify regions or genes having large effects on some traits (Table 1.2) (Rotschild 2003; Van Laere *et al.,* 2003; Dekkers 2004; Rotschild *et al.,* 2007; Muñoz *et al.,* 2009, Corominas *et al.,* 2013a; Ma *et al.,* 2014; Qiao *et al*., 2015)

Table 1.2. Strong candidate genes determining porcine production traits identified in QTL or GWAS analyses.

Trait	Candidate genes
Coat colour	*KIT, MC1R*
Growth, fatness and carcass composition	*IGF2, LEP, MC4R, MRF, PLAG1*
Meat quality	*CAST, ELOVL6, FABP4, FABP5, PHKG1, PRKAG3, RN, RYR1*
Litter size	*ESR, FSHB, PRLR, RBP4*
Disease susceptibility	*FUT1, NRAMP, SLA*

Despite the succes in the identification of genes containing causal mutations for some QTLs, their number is still very low due to the limitations of these approaches (Andersson & Georges, 2004). The main reasons may be: (1) production traits are complex and therefore difficult to detect, (2) genetic variants tend to explain a reduced amount of the genetic variation, and (3) several QTL studies are conducted in experimental crosses, where large QTL intervals have been identified due to linkage between markers. Moreover, several of the QTLs detected cannot be replicated maybe because they are segregating only in specific breeds or populations (Mackay *et al.,* 2009).

1.4.2. NGS technologies

The NGS platforms first released in 2005 offer to the scientific community the opportunity of exploring the whole genome, transcriptome or epigenome of an organism in a cost-effective manner (Morozova & Marra 2008). This approach allows to simultaneously sequence thousands-to-many-millions of fragments in parallel reactions (Yang *et al.,* 2014). To date, different systems (Roche 454, AB SOLiD, Illumina GA/HiSeq, Ion Torrent PGM, Illumina MiSeq) have been used for NGS of which Illumina HiSeq 2000 has the lowest reagent cost, the SOLiD has the highest accuracy and the Roche 454 has the longest read length (Soneson & Delorenzi 2013). A third generation sequencing method allowing the real-time sequencing of a single-molecule has been developed by *Pacific Bioscience* (Menlo Park, CA, USA) and has recently been used to improve the reference sequence of livestock species as cattle [Bovine Genome Project; http://www.hgsc.bcm.edu].

WGS provides an unprecedented opportunity to characterize the entire genome and to identify variants such as SNPs, indels and structural variants. In recent studies, WGS data has been employed to provide information of signatures of selection to understand the evolutionary history and speciation process (Amaral *et al.,* 2011; Rubin *et al.,* 2012; Ramírez *et al.,* 2014; Choi *et al.,* 2015; Moon et al., 2015; Paudel *et al.,* 2015) and study breed variability and homozygosity regions (Bosse *et al.,* 2012; Groenen *et al.,* 2012; Veroneze *et al.,* 2012; Esteve-Codina *et al.,* 2013; Ai *et al.,* 2015; Bianco *et al.,* 2015) by sequencing different *Sus* species and *Sus scrofa*

subespecies. However, the most promising application in livestock species is the use of WGS data to perform high resolution GWAS with the aim to detect variants and genes associated with complex diseases and its implementation in genomic selection (Bijma 2012; Pabinger *et al.,* 2013; Eynard *et al.,* 2015; Pérez-Enciso *et al.,* 2015). The basic workflow for the WGS analysis involves library preparation, sequencing, quality assessment, read alignment to the reference genome, variant calling, and results filtering.

On the other hand, in the transcriptomics field, the RNA-Seq approach provides information of the whole transcriptome including gene expression for each specific transcript in a large dynamic range, information of new genes and isoforms, splicing events, allele specific expression, and different promoter and polyadenylation signal usage in a single experiment (Wang *et al., 2*009). RNA-Seq experiments comprise: RNA isolation, library preparation, RNA sequencing, quality assessment and read alignment to the reference genome. The most common use of transcriptome profiling is the search for differentially expressed genes (Soneson & Delorenzi 2013). In this regard, different RNA-Seq studies have been performed in swine to identify differentially expressed genes and modulated pathways affected by sex (Esteve-Codina *et al.,* 2011), breeds (Sodhi *et al.,* 2014; Ghosh *et al.,* 2015), or important production traits (Chen *et al.,* 2011; Ramayo-Caldas *et al.,* 2012b; Corominas *et al.,* 2013b; Jiang *et al.,* 2013; Puig-Oliveras *et al.,* 2014a; Xing *et al.,* 2014); whereas, a few number of studies in pig have provided insights into SNVs, isoforms, and promoter and transcription start site usage (Jung *et al.,* 2012; Farajzadeh *et al.,* 2013) or alelle specific expression (Wu *et al.,* 2015). Therefore, these techniques are important tools for the identification of candidate genes for livestock production traits.

Different tools have been developed in the scientific community for NGS studies, including: raw data quality control and processing, alignment of the reads and visualization of data, short and large structural variant detection, and differential expression analysis (Table 1.3).

Table 1.3. Summary of bioinformatic tools for NGS data processing (Li & Homer 2010; García-Alcalde *et al.,* 2012; Martin *et al.,* 2012; Pabinger *et al.,* 2013; Zhao *et al.,* 2013). Tools used in this work are shown in bold characters.

Process	Tools
Quality assessment and read processing	ContEST, **FASTQC,** FASTX-Toolkit, Galaxy, htSeqTools, NGSQC, PIQA, PRINSEQ, **Qualimap**, SolexaQA, TagCleaner, TileQC
Alignment	BarraCUDA, BFAST, **Bowtie/Tophat,** BWA, ELAND, MAP, MAQ, Mosaik, mrsFAST, Novoalign, SMALT, SOAP, SSAHA, Stampy, YOABS
Variant calling	Atlas, Bambino, Beagle, CRIPSP, CoNAn-snv, CORTEX, Dindel, FreeBayes, GATK, IMPUTE2, Indelocator, Ion Variant Hunter, MaCH, moDIL, Pindel, PolyScan, Qcall, realSFS, SAMtools, Slinderll, Sniper, SNVer, SNVMix, SOAPindel, SOAPsnp, Syzygy, VarScan 2, VARiD, VipR
Structural variant identification	BreakDancer, CNAseq, CNVer, cnvHMM, CNVnator, CNV-seq, CONDEX, CoNIFER, CONTRA, ControlFreec, CopySeq, ExoCNVTest, ExomeCNV, ExomeDepth, PropSeq, RDXplorer, readDepth, Segseq, SeqGene, SVDetect, VarScan2, XHMM
Mapping and variant visualization	ABrowse, AnnoJ, Apollo, ARGO/Combo, Artemis, Bambino, BamView, Consed, DiProGB, EagleView, Ensembl, Gaggle, Gap5, GBrowse, G-compass, Genome Environment Browser, GenomeView, GenoViewer, Hawkeye, Integrated Genome Browser, **Integrative Genome Viewer,** JalView, JBrowse, LookSeq, MagicViewer, MapView, NGSView, SAMSCOPE, samtools tview, Savant, SeqWord, SNUGB, Tablet, UCSC genome browser, UTGB toolkit, VEGA, Vista
Differential gene expression	ShrinkSeq, **DESeq, edgeR,** NBPSeq, TSPM, voom, vst, baySeq, EBSeq, SAMseq

Alternatively, user friendly commercial tools which give the opportunity to perform all these analysis in a single platform, such as CLC bio (http://www.clcbio.com; CLC inc, Aarhus, Denmark), Galaxy (Giardine *et al.,* 2005; http://g2.bx.psu.edu) and Partek (http://www.partek.com; Partek Inc, St Louis, MO, USA) are available.

For the differential expression analysis of RNA-Seq data, the programs differ in a number of statistical issues, where the normalization is the main factor influencing the results (Hitzemann *et al.,* 2013). Although there is no consensus in the scientific community regarding which method performs best, Soneson & Delorenzi (2013)

reported that DESeq, EdgeR and NBPSeq give, in general, the best results for small sample size, and that DESeq is very conservative. Moreover, DESeq and edgeR show a slight reduction in type I error (the incorrect rejection of a true null hypothesis) rate when increasing sample size.

1.4.3. eQTL mapping approach

Transcriptomic data can be used to identify candidate genes underlying QTLs through the co-localization with eQTLs. This approach relies on the assumption that causative genes may have polymorphisms producing differences in their level of expression that translates into varying amounts of its corresponding functional protein and an observable phenotype difference (Verdugo *et al.,* 2010). To date, few eQTL studies have been performed to study porcine production traits (Table 1.4).

Table 1.4. eQTL studies for porcine production traits.

Trait	Reference
Growth	Steibel *et al.,* 2011; Ponsuksili *et al.,* 2012
Fatness	Steibel *et al.,* 2011; Cánovas *et al.,* 2012
Meat quality	Ponsuksili *et al.,* 2008, Ponsuksili *et al.,* 2010; Wimmers *et al.,* 2010; Heidt *et al.,* 2013; Ponsuksili *et al.,* 2014
Blood metabolite profile	Chen *et al.,* 2013a
Fatty acid composition	Muñoz *et al.,* 2013
Disease susceptibility and disorders	Liaubet *et al.,* 2011; Reiner *et al.,* 2014

The eQTL technique allows the mapping of transcript profiles as quantitative traits and to classify them in *cis* and *trans*-acting mode of action, identifying *hotspot* loci and regulators responsible for the phenotypes. By definition, *cis*-eQTLs are genetic variants mapped very close or into the studied gene that directly affect its level of transcription. In contrast, *trans*-eQTLs result from mutations in a different gene that may exert regulatory functions on the transcription of other genes (Doss *et al.,*

2005). Finally, a *hotspot* locus is defined as a region regulating the expression of a large number of genes.

1.4.4. Multi-trait network analyses for complex traits

Many livestock production traits are influenced by several genetic and environmental factors (Andersson 2001). Understanding the genetic basis determining these complex traits with the tools available to date is supposing a hard and arduous task. Several production traits are correlated and this can be due to pleiotropy or linkage (Chen & Lübberstedt 2010; Solovieff *et al.,* 2013) (Figure 1.8). Linkage occurs when two genes controlling different traits are located very close to each other, with little recombination rates, and therefore are transmitted together to progeny (Figure 1.8.B). A clear example of linkage in livestock occurred during the past years when applying a strong selection for higher lean meat content in pigs. In these populations, a recessive mutation in the *ryanodine receptor 1 (Skeletal)* (*RYR1*) gene causing malignant hyperthermia that mapped close to this locus rapidly increased its frequency resulting in low quality meats and also causing the death in animals under stress conditions (Andersson 2001). Pleiotropy occurs when a gene affects to more than one trait (Figure 1.8.A; Wagner & Zhang 2011) and has been described in growth, conformation and fat-related traits in pigs (Nagamine *et al.,* 2009; Fernández *et al.,* 2012; Ramayo-Caldas *et al*., 2012a; Muñoz *et al.,* 2013; Revilla *et al.,* 2014). A well-known case of pleiotropy in pigs is caused by the *v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog* (*KIT*) gene, which can affect coat colour, but also can exhibit lethal alleles, can affect the development of melanocytes, hematopoietic cells, primordial germ cells, interstitial cells in the small intestine and may affect hearing (Marklund *et al.,* 1998). Another clear effect of pleiotropy in pigs was evidenced by GWAS analyses for intramuscular fatty acid deposition and composition traits in Ramayo-Caldas *et al.* (2012a). In this study the authors described a pleiotropic effect for several intramuscular fatty acid composition indices in *Sus scrofa* chromosome 4 (SSC4), SSC8 and SSC16.

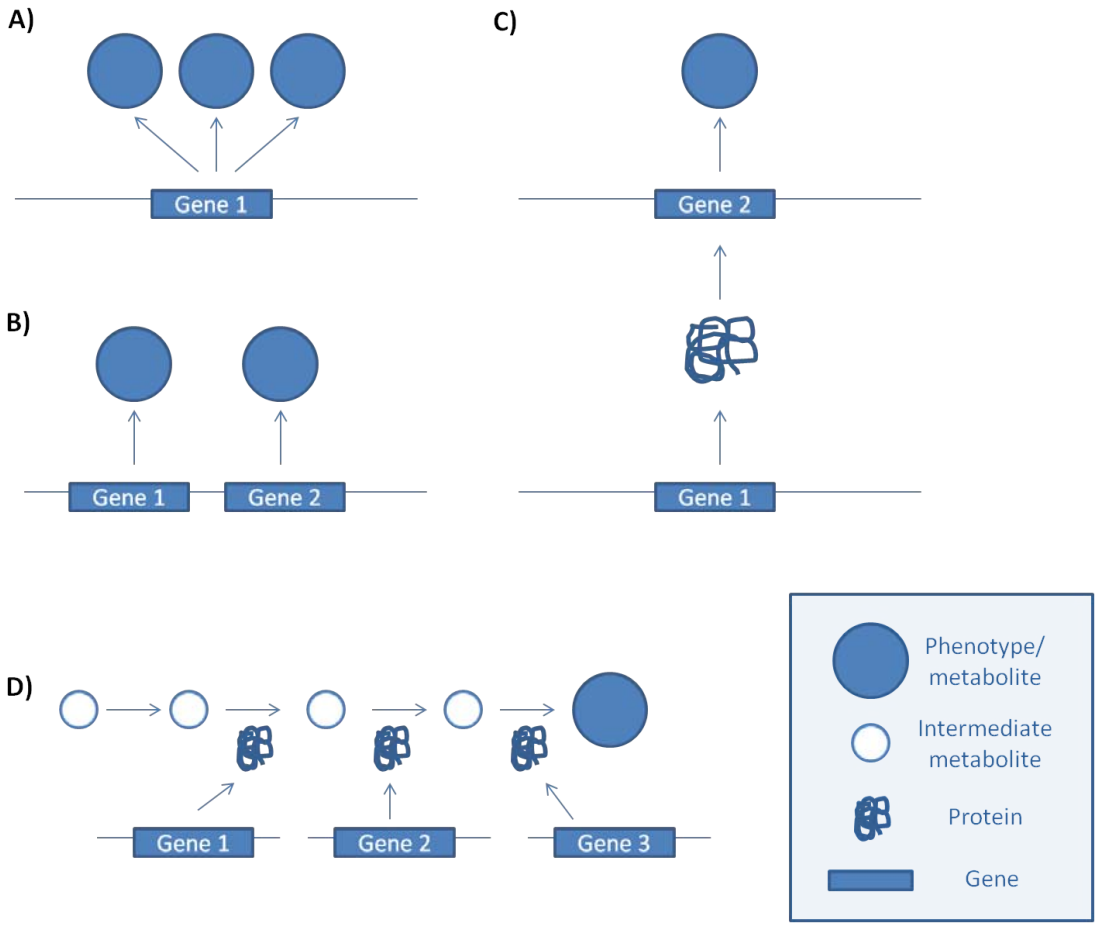


Figure 1.8. Schematic representation for possible causes of correlation among phenotypic traits (adapted from Wagner & Zhang 2011). A) Pleiotropy: a locus/gene affecting one more than one trait; B) Linkage: close genes affecting different traits that are inherited together due to linkage disequilibrium and thus causing differences in more than one phenotype; C) Gene-gene interactions determining phenotypes; D) Phenotype composed for different metabolites as intermediates in a single pathway.

A further factor determining phenotypic correlations among traits, an a cause of pleiotropy, is because genes do not act as independent units, they interact with each other in biologic complex networks, so a variation in one gene can alter a whole pathway (Figure 1.8.C). Finally, but not less important, the final phenotype can be viewed as the sum of all the reactions catalyzed for several enzymes playing different roles on the same pathway (Figure 1.8.D). This knowledge can be very useful to assess complex traits that are not independent of each other, for instance, the growth traits, where the weight of each primary cut will be more or less

proportional to the carcass weight. Correlations among traits must be considered in assessing the total impact of selection for a trait. As a trait is altered, performance for all correlated traits changes as well. Thus, trait correlation must be exploited in selection programs considering several traits at once in multi-trait analyses.

Multi-trait network approaches have recently been developed to exploit genetic correlations between traits (Bolormaa *et al.,* 2014). In this context, Fortes *et al.* (2010) developed a network approach based on co-association, named the *Association Weight Matrix* (AWM). The AWM is a systems biology approach that combines GWAS results from different correlated traits in an AWM for performing a partial correlation coefficient with information theory (PCIT). The AWM algorithm allows: (1) to capture SNPs in one trait of interest or in a minimum of three correlated traits having a relaxed statistical threshold ($p \leq 0.05$) in the GWAS, (2) to annotate the SNPs in genes, (3) to identify gene-gene interactions by performing correlations of the SNP additive effect across traits and (4) to include experimental knowledge about transcription factors and target interactions (Fortes *et al.,* 2010). Once identified all genes and their interactions, PCIT algorithm is used to assess significance threshold for interactions between genes (Reverter & Chan 2008). The obtained result is a gene-gene interaction network integrating positional genomic information of the SNPs and experimental knowledge of potential regulators. This approach assumes that genes with strongly correlated additive effects on a complex trait are likely to share genetic regulation. This approach has already been applied in a previous study of our group with the aim to gain insight into the molecular mechanisms determining the intramuscular fatty acid composition in pigs (Ramayo-Caldas *et al.,* 2014a). In this study, they predicted a network of 1,096 genes related to intramuscular fatty acid composition in pigs and suggested that genetic variants in three key transcription factors (*EP300*, *FHL2*, and *NCOA2*) were modulating the lipid metabolism and controlling the energy homeostasis in pigs (Ramayo-Caldas *et al.,* 2014a).

1.5. IBMAP cross animal material

The IBMAP_consortium (1996), a collaboration between the UAB (*Universitat Autònoma de Barcelona*), INIA (*Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria,* Madrid), and IRTA (*Institut de Recerca i Tecnologia Agroalimentàries,* Lleida) made possible the generation of the Iberian × Landrace cross (Figure 1.9).

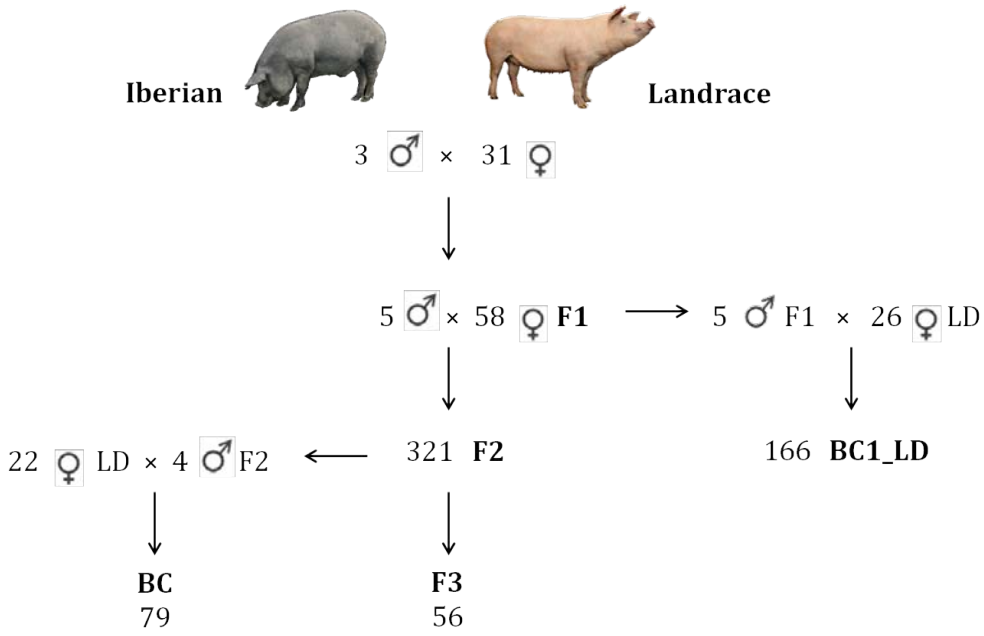


Figure 1.9. Schematic representation of the IBMAP cross (Iberian × Landrace).

This experimental cross was created for the identification of QTLs of interest for the porcine meat industry. The Iberian and the Landrace breeds were selected for being divergent pig lines basically differing in meat quality, growth, fatness, fertility, and feed efficiency traits. For instance, Iberian is a Mediterranean local breed from the Iberian Peninsula appreciated for its excellent meat quality and cured products with a higher content of SFA and MUFA fatty acids. Specifically, the Guadyerbas line is an original strain of the Iberian breed isolated in an experimental farm since 1945 (Toro *et al.,* 2000). In comparison, Landrace is a lean international breed that has undergone a strong selection for production benefits having a high prolificacy and growth. However, the Landrace meat is not as appreciated as the Iberian, having less intramuscular and higher content of PUFA (Serra *et al.,* 1998).

The IBMAP population was initially created by crossing 3 Iberian Guadyerbas boars (Dehesón del Encinar, Toledo) with 31 Landrace sows (Nova Genètica, Lleida), obtaining an F_1 generation from which different crosses were further generated: 321 F_2 animals obtained by mating F_1 animals, 56 F_3 individuals obtained by mating F_2 animals, 166 BC1_LD animals (25% Iberian × 75% Landrace backcross) from backcrossing five F_1 males with 26 Landrace sows, and 79 BC animals obtained by crossing four F_2 boars and 22 Landrace sows. All animals were fed *ad libitum* and sacrificed at 180 ± 2.8 days (average ± standard deviation) in a commercial slaughterhouse following national and institutional guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA, Institut de Recerca i Tecnologia Agroalimentàries).

1.5.1. Identification of QTLs in the IBMAP cross

QTL studies based on microsatellite markers in the IBMAP cross identified significant associated regions for growth, fatness and fatty acid composition on SSC2, SSC4, SSC6, SSC7, SSC8, and SSCX (Óvilo *et al.,* 2000; Pérez-Enciso *et al.,* 2000; Óvilo *et al.,* 2002a; Pérez-Enciso *et al.,* 2002; Varona *et al.,* 2002; Clop *et al.,* 2003; Mercadé *et al.,* 2005a; Óvilo *et al.,* 2005; Pérez-Enciso *et al.,* 2005). Afterwards, QTL and GWAS approaches using SNP markers from the PorcineSNP60 BeadChip of *Illumina* (Ramos *et al.,* 2009) have been employed, obtaining in general a higher resolution in QTL invervals and the identification of new genomic regions associated with the analysed traits (Fernández *et al.,* 2012; Ramayo-Caldas *et al.,* 2012a; Corominas *et al.,* 2013a; Muñoz *et al.,* 2013; Revilla *et al.,* 2014).

Within these QTLs several functional candidate genes have been identified for growth, fatness and meat quality traits including *IGF2, DECR, DGAT1, FABP2, FABP3,* ***FABP4, FABP5,*** *LEPR, ACADM, CDS1, CDS2, FABP2, MTTP, ELOVL6, FASN, GIP, ACACA,* and *ACSL4* (Clop *et al.,* 2002; Óvilo *et al.,* 2002b; Estellé *et al.,* 2005a; Estellé *et al.,* 2005b; Mercadé *et al.,* 2005a; Mercadé *et al.,* 2005b; Mercadé *et al.,* 2005c; Óvilo *et al.,* 2005; Estellé *et al.,* 2006; Kim *et al.,* 2006; Mercadé *et al.,* 2006a; Mercadé *et al.,* 2006b; Mercadé *et al.,* 2007; Estellé *et al.,* 2009a; Estellé *et al.,* 2009b; Muñoz *et al.,*

2009; Corominas *et al.,* 2012; Corominas *et al.,* 2013a) (genes analyzed in this work are shown in bold characters).

1.5.1.1. *FABP4* and *FABP5* candidate genes in SSC4

The first QTL in pigs, named FAT1, was found on SSC4 using a wild-boar intercross (Andersson *et al.,* 1994). Comparative maps identified this region to be homologous to both human chromosome 1 and 8. This QTL was subsequently replicated and validated in other crosses and other breeds (Knott *et al.* 1998; Marklund *et al.,* 1999; Walling *et al.,* 2000; Bidanel *et al.,* 2001; Cepica *et al.,* 2003; Sławińska *et al.,* 2009), including the IBMAP population (Figure 1.10) (Pérez-Enciso *et al.,* 2000). However, in IBMAP an additional QTL region was identified for growth and fatness traits (Mercadé *et al*., 2006b; Estellé *et al*., 2006) proximal to *FABP4* and *FABP5* genes.

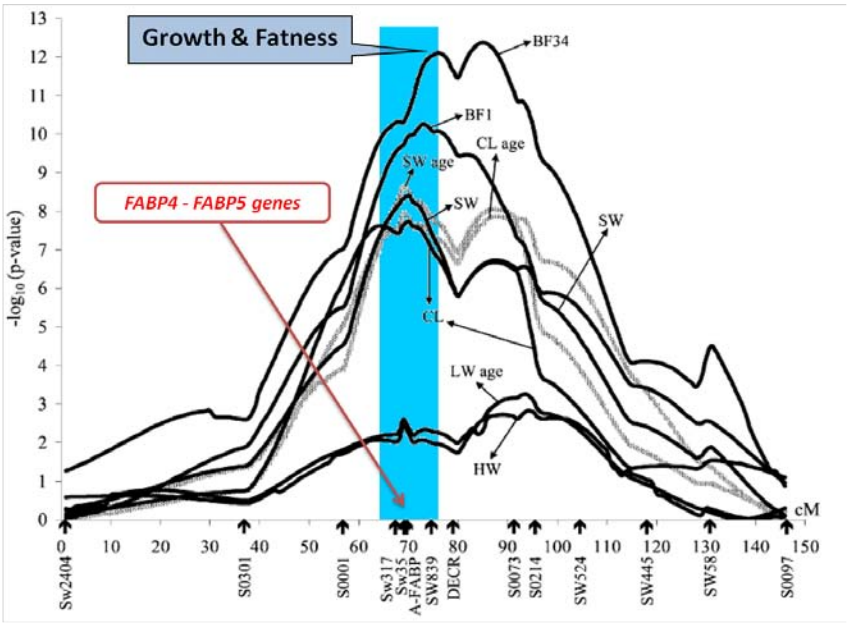


Figure 1.10. QTLs in SSC4 for growth and fatness traits detected in the IBMAP population (adapted from Mercadé *et al.,* 2005a).

In SSC4, *Fatty acid binding proteins 4* and *5* (*FABP4* and *FABP5*) genes have long been studied as candidate genes affecting fat deposition and composition traits in pigs (Gerbens *et al.*, 1998; Gerbens *et al.,* 2000; Gerbens *et al.,* 2001; Estellé *et al.,* 2006; Mercadé *et al.,* 2006b; Gao *et al.,* 2011; Chen *et al.,* 2013b) as well as in cattle, chicken and sheep (Hoashi *et al.,* 2008; Wang *et al.,* 2009b; Xu *et al.,* 2011). *FABP4*

and *FABP5* genes are members of the FABP family and are involved in the intracellular lipid transport of free fatty acids. The *FABP4* gene is highly expressed in adipocytes and it is also expressed in macrophages, dendritic cells, and skeletal muscle; meanwhile, *FABP5* gene is mainly expressed in epidermal cells, mammary gland, brain stomach, intestine, liver, kidney, testis, spleen, and placenta (Smathers & Petersen 2011). Furthermore, in previous studies of our group, two polymorphisms of *FABP4* (*g.2634_2635insC* in intron 1) and *FABP5* (*g.3000T>G* in intron 2) genes were genotyped in the F_2, F_3, and backcross generations of the IBMAP cross, revealing a tight association with fatness traits (Estellé *et al.,* 2006; Mercadé *et al.,* 2006b). Other studies of our group identified in the same region a pleiotropic QTL for backfat fatty acid composition traits in the IBMAP F_2 generation (Pérez-Enciso *et al.,* 2000; Clop *et al.,* 2003) and intramuscular fatty acid composition in the BC1_LD (Ramayo-Caldas *et al.,* 2012a).

2. OBJECTIVES

2. OBJECTIVES

This PhD thesis was done under the framework of the IBMAP Project funded by two projects: AGL2008-04818-C03/GAN (MICINN) and AGL2011-29821-C02 (MINECO). The animal material generated by the IBMAP project consisting in an Iberian × Landrace cross was obtained thanks to the collaboration between INIA, IRTA and UAB research groups.

The main aim of this thesis was to increase the knowledge of the genetic basis determining growth and fat-related traits in pigs.

The specific objectives were:

1) To evaluate the porcine *FABP4* and *FABP5* as candidate genes for fatty acid composition traits and study their expression and regulation.

2) To characterize the transcriptome architecture of the porcine *Longissimus dorsi* muscle and to identify genes and pathways determining the differences in the intramuscular fatty acid composition among animals.

3) To study gene interactions, pathways, and main regulators determining growth and fat-related traits by the construction of a co-association gene network.

4) To study the expression and to identify eQTLs of a set of 45 selected candidate genes for fat content and fatty acid composition in muscle.

3. PAPERS AND STUDIES

Analysis of *FABP4* and *FABP5* gene expression and polymorphisms affecting pig fatty acid composition

Puig-Oliveras A, Ballester M, Castelló A, Gago-Díaz M, Fernández AI, Folch JM

Genetics Selection Evolution. In revision

Analysis of *FABP4* and *FABP5* gene expression and polymorphisms affecting pig fatty acid composition

Anna Puig-Oliveras[1,2 §], Maria Ballester[2,3,4,5], Anna Castelló[1,2], Marina Gago-Díaz[1], Ana I Fernández[6], and Josep M Folch[1,2]

1 Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain, **2** Plant and Animal Genomics, Centre de Recerca en Agrigenòmica (CRAG), 08193 Bellaterra, Spain, **3** Génétique Animale et Biologie Intégrative UMR1313 (GABI), Institut National de la Recherche Agronomique (INRA), 78350 Jouy-en-Josas, France, **4** Génétique Animale et Biologie Intégrative UMR1313 (GABI), AgroParisTech, 78350 Jouy-en-Josas, France, **5** Laboratoire de Radiobiologie et Etude du Génome (LREG), Commissariat à l'énergie atomique et aux énergies alternatives (CEA), 78350 Jouy-en-Josas, France, **6** Departamento de Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040 Madrid, Spain.

[§] Corresponding author

Email addresses:

AP: anna.puig@cragenomica.es

MB: mballester@jouy.inra.fr

AC: anna.castello@cragenomica.es

MG: mgagodiaz@gmail.com

AIF: avila@inia.es

JMF: josepmaria.folch@uab.es

Abstract

Background: The *FABP4* and *FABP5* genes, coding for fatty-acid-transport proteins, have been studied as positional and functional candidate genes for SSC4 QTLs for fat and growth related traits. Polymorphisms in these genes have been associated with fat deposition and growth traits. More recently, QTLs affecting fatty acid composition in backfat and intramuscular fat have been identified in this genomic region. Thus, the aim of this study was to evaluate the *FABP4* and *FABP5* genes as positional and functional genes affecting fatty acid composition in muscle and backfat tissues.

Results: The association analysis with *FABP4* and *FABP5* gene polymorphisms and the QTL scan performed in the Iberian × Landrace backcross identified the *FABP4:g.2634_2635insC* polymorphism as the most significant marker for palmitoleic and eicosatrienoic fatty acids and PUFA content in muscle. The *FABP5:g.3000T>G* polymorphism was the most associated marker for oleic and linoleic fatty acids and MUFA content in muscle. Furthermore, the *FABP4* and *FABP5* gene expression analysis in 114 BC1_LD animals revealed that *FABP4* gene expression in backfat, but not in muscle, was associated with *FABP4:g.2634_2635insC*. In contrast, *FABP5:g.3000T>G* was not associated with gene expression levels. The eGWAS highlighted the *FABP4:g.2634_2635insC* polymorphism as the most associated polymorphism with *FABP4* gene expression in backfat while different genomic regions were associated in *trans* with *FABP5* gene expression in muscle. Finally, a putative transcription binding site for PPARG may be affected by *FABP4:g.2634_2635insC* polymorphism, modifying *FABP4* gene expression and determining the intramuscular fatty acid composition.

Conclusions: Our results suggest *FABP4* as an important gene regulating fatty acid composition, being *FABP4:g.2634_2635insC* polymorphism the strongest signal associated both to *FABP4* gene expression in backfat and to intramuscular fatty acid composition in muscle.

Background

The genetic basis of fat composition has largely been studied in swine due to their impact in meat production yield and quality. Fatty acid (FA) composition can affect the taste and the quality of the cooked and the cured meat [1,2]. Specifically, high polyunsaturated FA (PUFA) content negatively affects the oxidative stability, flavour and colour of pork meat. Moreover, the meat nutritional value can also be altered by the FA profile, influencing human health. For instance, the n-3 PUFA are considered to reduce cholesterol concentration and decrease inflammatory processes [3,4]. Thus, meat with higher amounts of intramuscular fat (IMF), a healthier FA profile, and good organoleptic properties may have better consumer acceptance. Hence, there is an interest for the pig industry to produce meat with a balanced FA composition [5].

Over the last years, an Iberian × Landrace intercross (IBMAP) was generated for the identification of QTLs affecting meat and carcass quality traits, such as growth, fatness, and FA composition in backfat (BF) and IMF [6-11]. These breeds are divergent for the studied traits, for instance, Iberian is a local an obese Mediterranean breed with a higher content of SFA and MUFA, specifically palmitic (C16:0) and oleic (C18:1(n-9)) FAs; whereas Landrace is a lean international breed with a higher content of PUFA such as linoleic (C18:2(n-6)) and α-linolenic (C18:3(n-3)) [1]. In the F_2 generation of this cross, a pleiotropic QTL effect was described on SSC4 for the percentage of C18:1(n-9) and C18:2(n-6) FAs and the double-bond (DBI) and peroxidability (PI) indexes in BF [6,7]. QTLs for C18:2(n-6) FA content in BF, perirenal fat and abdominal fat were reported in the same region in other swine populations [12-14]. More recently, a genome-wide association study (GWAS) has identified the same SSC4 region as associated with the intramuscular percentage of palmitoleic (C16:1(n-7)), C18:1(n-9), and C18:2(n-6) FAs in an Iberian × Landrace backcross (BC1_LD) [10]. Besides, a QTL scan performed in the BC1_LD detected a significant QTL on SSC4 for IMF content of C16:1(n-7), C18:2(n-6), and eicosatrienoic (C20:3(n-6)) FAs and backfat content of C16:0, C18:2(n-6), C18:3(n-6), and C20:3(n-6) FAs [11].

Fatty acid binding proteins 4 and *5* (*FABP4* and *FABP5*) genes have long been studied as candidate genes affecting fat-related traits in pigs [15-20]. FABPs are responsible for

the transport of long chain FAs for different purposes: uptake, storage, and utilization [21]. They are also involved in other processes such as adipogenesis, myogenesis, and regulation of gene expression [22,23].

In previous studies of our group, *FABP4* and *FABP5* genes were sequenced in the IBMAP founding generation (three Iberian boars and seven Landrace sows) to identify polymorphisms. *FABP4*:*g.2634_2635insC* and *FABP5*:*g.3000T>G* genetic variants were genotyped in the F_2, F_3, and BC1_LD generations of the IBMAP cross, revealing a tight association with fatness traits [18,19]. The aim of the present study was to evaluate *FABP4* and *FABP5* as candidate genes for the SSC4 QTL affecting FA composition in IMF and BF in an Iberian × Landrace backcross. To achieve this goal, gene expression analysis and association studies between *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms with FA composition in BF and muscle (*Longissimus dorsi*) were performed.

Methods

Animal samples and phenotypic records

The animal material used in this study belongs to the IBMAP cross population generated by the initial crossing of three Iberian (Guadyerbas line) boars with 31 Landrace sows [6], and containing several generations and related backcrosses. Here, we have analyzed 144 animals from a backcross generation (BC1_LD; 25% Iberian × 75% Landrace) produced by crossing five F_1 (Iberian × Landrace) boars with 26 Landrace sows. All animals were fed *ad libitum* with a cereal-based commercial diet. Pigs were slaughtered at an average age of 179.8 ± 2.3 days following national and institutional guidelines for the ethical use and treatment of animals in experiments and approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries). Samples of BF (taken between the third and the fourth ribs) and muscle (*Longissimus dorsi*) were collected, snap-frozen in liquid nitrogen, and stored at -80ºC for RNA extraction. DNA was extracted from blood samples and used for gene PCR amplification and polymorphism genotyping.

Muscle and BF FA composition was determined as previously described [10,11]. The FA measures in IMF and BF were: myristic (C14:0), palmitic (C16:0), palmitoleic (C16:1(n-7)), heptadecanoic (C17:0), heptadecenoic (C17:1), stearic (C18:0), octadecenoic (C18:1(n-7)), oleic (C18:1(n-9)), linoleic (C18:2(n-6)), α-linolenic (C18:3(n-3)), arachidic (C20:0), eicosenoic (C20:1(n-9)), eicosatrienoic (C20:3(n-6)), arachidonic (C20:4(n-6)), polyunsaturated (PUFA), monounsaturated (MUFA), and saturated (SFA) FAs.

Genotyping data

A total of 179 animals from the BC1_LD (144 backcrossed individuals and their corresponding 35 parents) were genotyped with the Porcine *SNP60 Beadchip* (Illumina) following the Infinium HD Assay Ultra protocol (Illumina) [24]. Raw data was visualized with GenomeStudio software (Illumina) and trimmed for high genotyping quality (call rate > 0.99). Plink [25] software was used to retain markers with a minor allele frequency (MAF) ≥ 5% and animals with missing genotypes < 5 %. After the quality control filter, a subset of 40,476 SNPs and 179 animals remained. The position of the SNPs was based on the Sscrofa10.2 assembly [26]. Furthermore, all animals (n=179) were also genotyped for the *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms following the pyrosequencing protocols described by Mercadé *et al.* [19] and Estellé *et al.* [18], respectively.

Identification of putative binding sites and its conservation across species

A computer-assisted identification of putative transcription binding sites in intron 1 of *FABP4* was performed using MatInspector (cut-off as default) [27] from GenomatixSuite software (Genomatix Software GmbH) with the Genomatix Matrix Library 8.3. To explore the conservation of the *FABP4* exon 1 and proximal intron 1 sequence region harboring *FABP4:g.2634_2635insC* polymorphism, *FABP4* gene sequences for the human (NC_000008), cow (AC_000171) and pig (Y16039) were downloaded from NCBI and compared by multi-alignment analysis using Multalin program [28].

QTL scan analysis

In order to select properly located markers for QTL scan analysis, the linkage map of SSC4 of the IBMAP population [29] was used. A total of 155 genetic markers were chosen for further analyses according to their higher informativity (I_e > 0.6) in the BC1_LD calculated by the Ron index [30]. In addition, markers *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* were included in the association analyses. For this subset of 157 SNPs, genetic distances were calculated for the BC1_LD population using the "Fixed" option of CRI-MAP program v2.503 [31]. From the clusters of markers belonging to the same linkage group, without recombination between them, only the most informative SNP was selected, remaining a total of 78 SNPs [see Additional file 1: Table S1].

QTL scan was performed for FA composition in IMF and BF by using the basic model implemented on Qxpak 5.0 [32]:

$$y_{ijk} = Sex_i + Batch_j + \beta c_k + P_{ak} + u_k + e_{ijk,} \qquad (1)$$

in which y_{ijk} is the k^{th} individual's record observation for the analyzed trait, depending on fixed effects for sex and batch (with two and five levels, respectively), carcass weight as a covariate and its respective slope (βc), a is the QTL additive effect, P_{ak} is the additive coefficient calculated as $P_{ak} = Pr(QQ) - Pr(qq)$, the probability of the k^{th} individual being homozygous for alleles of Iberian origin minus the probability of being homozygous for alleles of Landrace origin, u_k represents the infinitesimal genetic effect with random distribution $N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ where **A** is the numerator of the pedigree-based relationship matrix and e_{ijk} the random residual.

To assess the statistical significance a Bonferroni correction was calculated with the R package [33].

Genotype association analysis

Association analyses with the polymorphisms *FABP4 g.2634_2635insC* and *FABP5:g.3000T>G* and FA composition in BF and IMF were performed using a mixed model implemented in Qxpak 5.0 [32]:

$$y_{ijkl} = Sex_i + Batch_j + \beta c_k + \lambda_k a_l + u_k + e_{ijkl}, \qquad (2)$$

in which y_{ijkl} is the k^{th} individual's record, sex (two levels) and batch (five levels) are fixed effects, β is a covariate coefficient with *c* being carcass weight, λ_k is a -1, 0, +1 indicator variable depending on the k^{th} individual's genotype for the l^{th} SNP, a_l represents the additive effect associated with the l^{th} SNP, u_k is the infinitesimal genetic effect with random distribution $N(\mathbf{0}, \mathbf{A}\sigma_u^2)$, where **A** is the numerator of the pedigree-based relationship matrix and e_{ijkl} the residual.

Genomic intervals of ±1Mb around the most significant SNPs were annotated using BIOMART [34].

RNA isolation and gene expression quantification

Total RNA was isolated from muscle and BF tissues of 114 BC1_LD animals using the RiboPure™ Isolation of High Quality Total RNA (Ambion) and quantified in a NanoDrop ND-1000 spectrophotometer (NanoDrop products). Once isolated, the cDNA was synthesized *de novo* from total RNA (1 µg from muscle or 0.3 µg from BF) in a 20 µl mix using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems) and random hexamers primers. Real time quantitative PCR (RT-qPCR) primers are shown in Table S2 [see Additional file 2: Table S2]. Three genes were analyzed as endogenous controls by GeNorm [35]: *β-2 microglobulin* (*β2M*), *Hypoxanthine phosphoribosyltransferase 1* (*HPRT1*) and *Glyceraldehyde 3-phosphate dehydrogenase* (*GADPH*). *HPRT1* and *β2M* were chosen as the best endogenous controls for both tissues. PCR amplification was performed in triplicate in a 20-µl final volume using FastStart Universal SYBR Green Master (Rox; Roche Applied Biosystems) and containing 5 µl of cDNA diluted 1:125 from muscle and BF. Primer concentration was at 600 nM for *FABP4*, *FABP5* and *β2M* and 900 nM for *HPRT1* in both tissues. PCR amplification was performed using 96-well optical plates in an ABI PRISM 7900HT sequence Detection System (Applied Biosystems) under the following conditions: 10 min at 95 ºC, 40 cycles of 15 s at 95 ºC and 1 min at 60 ºC. To assess primer-dimer formation, a dissociation curve was drawn for each primer pair. To quantify and normalize the relative quantification (RQ) data, the $2^{-\Delta\Delta CT}$ method [36] was applied using a sample with low expression as a calibrator. R software [33] was employed for statistical comparison of gene expression data and *FABP4* and *FABP5* genotypes using

a linear procedure considering sex and batch as fixed effects. Differences were considered as statistically significant at p-value < 0.01.

Gene expression association analysis (eGWAS)

Association analyses of whole-genome SNP genotypes and *FABP4* and *FABP5* mRNA expression values from muscle and BF were also performed. A total of 40,476 SNPs mapping to a proper location in the Sscrofa10.2 assembly and *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms were used. For *FABP5* gene, backfat and muscle mRNA expression data did not fit a normal distribution and the RQ value was transformed to $\log_2$. The previously described model (2) without the carcass weight covariate was used for the expression genome-wide association studies (eGWAS). To measure the statistical significance at the genome-wide level for association studies, the R package q-value [33,37] was used to calculate the false discovery rate (FDR)-based q-value (set at q-value ≤ 0.1). Genomic intervals of ±1Mb around the most significant SNP were annotated using BIOMART [34].

Results

Effect of *FABP4* and *FABP5* gene polymorphisms on fatty acid composition

In the present study, a total of 179 animals belonging to the BC1_LD IBMAP population (144 backcrossed individuals and their corresponding parents) were genotyped for the *FABP4*:*g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms.

In order to test the association of *FABP4* and *FABP5* polymorphisms with FA composition in muscle (n=125) and BF (n=130), an association analysis using an additive model (1) was performed in BC1_LD animals. Significant associations were observed between polymorphism *FABP4*:*g.2634_2635insC* and the IMF content of C16:0, C16:1(n-7), C18:1(n-7), C18:2(n-6), and C20:3(n-6) FAs, MUFA and PUFA (Table 1). Homozygous individuals for the *FABP4*:*g.2635insC* allele, which is fixed in Iberian boars and is found at low frequency (0.23) in Landrace sows, showed a higher IMF content of C16:0, C16:1(n-7), and C18:1(n-7) FAs and MUFA and a decreased IMF content of C18:2(n-6) and C20:3(n-6) FAs and PUFA (Figure 1A). Besides, in backfat,

polymorphism *FABP4:g.2634_2635insC* showed a significant association with the percentages of C16:0, C18:2(n-6), C18:3(n-3), and C20:3(n-6) FAs and PUFA (Table 1). *FABP4*:*g.2635insC* homozygous individuals showed more C16:0 and less C18:2(n-6), C18:3(n-3), and C20:3(n-6) FAs and PUFA content in BF (Figure 1B).

Significant associations were also found between SNP *FABP5:g.3000T>G* and C18:1(n-9) and C18:2(n-6) FAs, MUFA and PUFA in IMF (Table 1). Finally, in backfat, significant associations were only observed between SNP *FABP5:g.3000T>G* and C18:1(n-9) FA and MUFA content (Table 1). Individuals that have inherited the *FABP5g.3000T* allele, which is fixed in Iberian boars and shows intermediate frequencies (0.42) in Landrace sows, showed in muscle a higher C18:1(n-9) FA and MUFA content, a decreased C18:2(n-6) FA content and suggestive lower PUFA levels. Conversely, these animals showed in BF a higher content of C18:1(n-9) and a suggestive increase in MUFA content (Figure 1C and 1D).

QTL scan on SSC4 for fatty acid composition in backfat and muscle

A QTL scan for the IMF and BF FA content was performed on the BC1_LD population to assess the association of the *FABP4* and *FABP5* chromosomal region with the analyzed traits. QTL analysis was performed in those traits significantly associated with *FABP4* and *FABP5* polymorphisms (Table 1, traits with p-values < 0.01).

FABP4 gene remains un-annotated in the current pig genome reference sequence *Sscrofa10.2*. RH and linkage maps located *FABP4* gene close to *FABP5* gene on SSC4 [18,38], in agreement with the human comparative map. Accordingly, *FABP4* gene was located at 1 cM before *FABP5* in the linkage map for further analysis [see Additional file 1: Table S1] [18].

A total of 78 SNPs from the 60K SNP pig chip (Illumina) were selected on the basis of their informativity [30] and position in the linkage map [29]. *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms were also included in this analysis [see Additional file 1: Table S1].

A QTL for IMF content of C16:1(n-7), C18:1(n-9), C18:2(n-6), and C20:3(n-6) FAs, MUFA and PUFA was located around 50 cM on the linkage map where *FABP4* and *FABP5* genes were mapped [see Additional file 3: Figure S1]. Furthermore, the *FABP4-FABP5* genomic region was significantly associated with C18:2(n-6) and PUFA content in BF [see Additional file 3: Figure S1], although the largest QTL effect was surrounding the 59-108 cM region [see Additional file 3: Figure S1]. For C16:0 and C18:1(n-7) FAs in IMF and C16:0, C18:1(n-9), C18:3(n-3), and C20:3 (n-6) FAs and MUFA in BF no significant QTL were found in the *FABP4-FABP5* genomic region. These results are in accordance with a previous GWAS performed in the same animal population, in which the *FABP4-FABP5* genomic region was associated with C16:1(n-7), C18:1(n-9), and C18:2(n-6) FA content in IMF [10]. A previous QTL scan performed by Muñoz *et al.* [11] in the same animal material only detected a significant QTL on the *FABP4-FABP5* genes chromosomal region (approximately at 44 cM in their linkage map) for C18:2(n-6) and C20:3(n-6) FAs in IMF. In addition to these two FAs, we report significant QTLs for C16:1(n-7) and C18:1(n-9) FAs, PUFA and MUFA content in IMF. Differences between both studies may be due to the number and selection of the markers. In the present study 78 SNPs of SSC4 were selected according to their informativity and position, whereas in Muñoz *et al.* [11] 8,417 SNPs from all autosomes were retained according to their position in the linkage map.

In order to test if polymorphisms *FABP4:g.2634_2635insC* or *FABP5:g.3000T>G* explained a major part of the QTL variability, an analysis containing both the QTL effects and the gene polymorphism effect (*FABP4:g.2634_2635insC* or *FABP5:g.3000T>G* effect) was performed. The QTL effect was tested removing it from the full model and the polymorphism effect was tested removing it from the full model. The results of likelihood ratio (LR) test showed that the *FABP4:g.2634_2635insC* or the *FABP5:g.3000T>G* polymorphism effect was more significant than the QTL effect for all analyzed traits except for C18:2(n-6) FA backfat content. Note that the QTL effect disappeared after fitting gene polymorphisms as a covariate in the model (Table 2).

Finally, for C18:2(n-6) FA and PUFA content in BF, models fitting one QTL against a model considering two different QTL were tested to determine whether one or two

QTL were segregating on SSC4. Results of the LR test indicated that the model with one QTL on 58-109 cM was the most likely for these two traits (LR $_{C18:2(n-6)}$ = 10.61; p-value $_{C18:2(n-6)}$ = 1.12×10^{-3}; LR $_{PUFA}$ = 10.31; p-value $_{PUFA}$ = 1.32×10^{-3}), concluding that no QTL was segregating for backfat FA composition in the *FABP4-FABP5* genomic region (LR $_{C18:2(n-6)}$ = 1.60; p-value $_{C18:2(n-6)}$ = 2.93×10^{-1}; LR $_{PUFA}$ = 1.10; p-value $_{PUFA}$ = 2.06×10^{-1}).

Study of *FABP4* and *FABP5* gene expression

Previous studies using microarrays have found significant differences in the expression of *FABP4* and *FABP5* genes in pigs with different profiles of fatty acid content in muscle [39,40]. Hence, with the aim to study the association between polymorphisms *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* and gene expression, the *FABP4* and *FABP5* expression profiles were analyzed by RT-qPCR in 114 BC1_LD animals in muscle and BF.

FABP4 gene expression level in adipose tissue was higher than in muscle (data not shown), in agreement with other studies reporting that *FABP4* is abundantly expressed in adipose tissue [41,42]. In contrast, *FABP5* is expressed at low levels in adipocytes and muscle tissues (data not shown).

A moderate correlation between *FABP4* and *FABP5* gene expression was observed in muscle (r = 0.45, p-value = 3.57×10^{-6}) and in BF (r = 0.54, p-value = 1.10×10^{-8}). Also, *FABP5* gene expression in muscle was higher in females than in males (p-value = 3.1×10^{-4}).

Finally, no clear correlation between *FABP4* and *FABP5* expression among tissues was observed, suggesting that different mechanisms controlling *FABP4* and *FABP5* expression operate in BF and muscle tissues.

FABP4 and *FABP5* gene expression values were classified according to *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* genotypes. A significant difference in *FABP4* gene expression among *FABP4:g.2634_2635insC* genotypes was observed in BF (p-value = 4.39×10^{-4}), but not in muscle (Figure 2). This result supports the hypothesis of different mechanisms controlling *FABPs* gene expression among tissues. Animals

with the *FABP4:g.2635insC* allele showed a higher *FABP4* expression in BF (p-value = 1.85×10^{-5}; â (estimated additive effect) = 0.139) in comparison with the *FABP4:g.2634_* allele (Figure 2). No significant differences were observed for *FABP5* gene expression in muscle or BF among *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* genotypes.

Genome-wide association studies of *FABP4* and *FABP5* gene expression

Taking into account that differences in *FABP4* and *FABP5* gene expression were observed among tissues, eGWAS (expression GWAS) were performed using 40,476 SNPs of the porcine 60 K SNP Chip (Illumina) and the *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms and the RQ expression data of *FABP4* and *FABP5* genes for BF and muscle.

In muscle, no significant association was found for *FABP4* gene expression (Figure 3A). Noteworthy, the most significant peak for *FABP4* gene expression in BF was located in SSC4 (Figure 3B; significant at chromosome-wise level) being *FABP4:g.2634_2635insC* the most associated polymorphism (p-value=4.71×10^{-6}).

For *FABP5* gene expression four potential *trans* acting regions in chromosomes SSC4, SSC6, SSC9, and SSC13 were identified (Figure 3C). In SSC4, three SNPs at position 100.7 Mb were identified as the most significantly associated (ALGA0026790, ALGA0026791, MARC0042655, p-value= 7.47×10^{-6}); two of them were located in the upstream and downstream regions of the *Fc receptor-like 1* (*FCRL1*) gene, an immunoglobulin receptor involved in inflammation pathways activated with high free fatty acids levels [43]. Another potential *trans* eQTL was identified in SSC6 (MARC0010912, 24.8 Mb, p-value = 3.82×10^{-7}) close to *E2F transcription factor 4, p107/p130-binding* (*E2F4*) gene, which acts as a repressor of *peroxisome-proliferator-activated receptor gamma* (*PPARG*), a well-known adipogenic transcription factor [44]. In SSC9, a significant region was found at 54.9 Mb (ASGA0082302, p-value = 1.61×10^{-6}), where the *CXADR-like membrane protein* (*CLMP*) gene was identified. This gene plays a role in adipocyte differentiation and development of obesity [45]. Finally, the most significant associated region found in SSC13 was at position 17.3 Mb (H3GA0035652, p-value = 1.81×10^{-5}). In this region, the *5-azacytidine induced 2* (*AZI2*) gene was annotated. *AZI2* gene contributes to the activation of the *nuclear factor of kappa light*

polypeptide gene enhancer in B-cells (*NFκB*) which is involved in lipid metabolism [46]. Whereas, no clear significant associated regions were observed for *FABP5* gene expression in BF (Figure 3D).

PPARG may be involved in _FABP4_ gene expression regulation

FABP4g.2634_2635insC polymorphism was clearly associated with *FABP4* gene expression and FA composition in muscle and backfat. This polymorphism was located in intron 1 of *FABP4* gene [19], in a region conserved between different species [see Additional file 4: Figure S2]. To assess if *FABP4:g.2634_2635insC* polymorphism could affect *FABP4* gene expression through the disruption of a transcription factor-binding site, an *in-silico* identification of *cis*-acting DNA-sequence motifs in this region was carried out. Remarkably, *FABP4:g.2634_2635insC* polymorphism was found to be in the binding site of the *PPARG* [see Additional file 4: Figure S2].

Discussion

Free FAs (FFAs) are transported through the cytoplasm by binding proteins like FABPs that deliver them to specific compartments in the cell, such as the mitochondria or the peroxisome for their oxidation; to the endoplasmic reticulum for signaling and membrane synthesis; to the lipid droplet for storage; or to the nucleus for activation of nuclear receptors (reviewed in [47]).

FABP4 is mainly secreted by adipocytes [48] and has been suggested to play an important role in FA storage in adipocytes as triacylglycerol [49]. During the initial steps of lipogenesis, C16:0 and C18:0 FAs are converted to the most abundant MUFAs in pork (C16:1 and C18:1) [50] which are important components of membrane phospholipids, cholesterol and triglycerides [51].

Taken together, the association with *FABP4* and *FABP5* gene polymorphisms and the classical QTL analysis showed that *FABP4*:*g.2634_2635insC* was the most significant marker for palmitoleic acid (C16:1(n-7)), C20:3(n-6) FA, and PUFA composition in *Longissimus dorsi* muscle. Remarkably, palmitoleic acid has been described as a marker for *de novo* FA lipogenesis because very little amount of this FA is provided by the diet [52]. In addition, *FABP5:g.3000T>G* was the SNP showing the strongest statistical

significance for oleic acid (C18:1(n-9)), linoleic acid (C18:2(n-6)), and MUFA percentages in muscle. Oleic acid is the most abundant and nutritionally relevant FA of pig meat and it is associated with the flavor of the Iberian pig dry-cured products. Linoleic acid is an essential FA which must be provided by the diet and it is the most abundant omega 6 FA of pig meat.

These results are in agreement with the reported phenotypic differences in fat deposition and composition between Iberian and Landrace pigs [1]. Iberian pigs have higher percentages of palmitic acid, oleic acid, SFA and MUFA, and lower concentrations of linoleic and α-linoleic acids than commercial breeds [1]. Here, animals homozygous for Iberian *FABP4g.2635insC* and *FABP5g.3000T* genotypes (more frequent in Iberian than in Landrace parental lines) showed a higher percentage of MUFA and lower content of PUFA both in muscle and backfat tissues.

The analysis of *FABP4* and *FABP5* gene expression in backfat and muscle revealed that *FABP4* gene expression in backfat, but not in muscle, was associated with polymorphism *FABP4g.2634_2635insC*. Although *FABP5g.3000T>G* SNP was not associated with gene expression levels, we cannot discard this gene as candidate for FA composition determination. Further studies are needed to identify new polymorphisms in the *FABP5* gene. Conversely, the eGWAS study pointed to the *FABP4g.2634_2635insC* polymorphism as the most strongly associated with *FABP4* gene expression in backfat whereas no clear peak was observed for *FABP4* gene expression in muscle. Hence, a concordant QTL and eQTL location was only observed for *FABP4* gene in adipose tissue and IMF FA composition traits, reinforcing *FABP4* as a strong candidate gene for meat quality traits. A more complex pattern was observed for *FABP5* gene expression in muscle with four *trans*-acting regions identified in chromosomes SSC4, SSC6, SSC9, and SSC13. The associated SNPs in SSC4 were located at 100.7 Mb, whereas *FABP5* gene was mapped at 60.3 Mb in the *Sscrofa10.2* assembly of the pig genome. Relevant positional candidate genes were identified in the *trans*-eGWAS regions including *FCRL1* (SSC4), *E2F4* (SSC6), *CLMP* (SSC9), and *AZI2* (SSC13). Further work is needed to unravel the role of these genes in the regulation of *FABP5* gene. In spite of the correlation observed between *FABP4* and *FABP5* genes in muscle and backfat, eGWAS results suggest that different mechanisms of genetic regulation

are acting on each gene. Moreover, considering a single gene, different patterns of *trans*-acting regions were observed in muscle and backfat.

The higher level of *FABP4* gene expression in BF for animals homozygous for the *FABP4g.2635insC* allele may positively impact the intracellular uptake of C16 and C18 for *de novo* lipogenesis of SFA and MUFA and for triacylglycerol storage. *FABP4* has been described to play a central role in obesity development [53,54], being *FABP4* knockout mice protected against diet-induced obesity [55,56]. In concordance, the Iberian phenotype shows higher IMF and strongest backfat tissue development, being more prone to obesity. This result agrees with a previous analysis of pig adipose tissue transcriptome in animals of the BC1_LD, in which a higher expression of lipogenic genes was observed in animals with a higher content of MUFA and SFA in IMF [57]. The high MUFA and SFA content positively affects the firmless/oilness of adipose tissue and the oxidative stability of meat [2].

Furthermore, FABP4 is also involved in the delivery of FFA to the muscle in order to fulfill the required energy [58]. Thus, the higher expression levels of *FABP4* in adipose tissue observed in animals inheriting the Iberian allele (*FABP4g.2635insC*) may be affecting the transport of FA to the muscle and therefore determining the differences in FA composition observed among Iberian and Landrace breeds [1]. This agrees with the impaired FA uptake on muscle tissue observed in $FABP4^{-/-}$ and $FABP5^{-/-}$ knockout mice [58]. In FABP-deficient mice, an increase in fatty acid oxidation and energy expenditure has been described [59]. This agrees with our previous RNA-Seq study in liver [60], adipose tissue [57], and muscle [61] in which extreme BC1_LD animals for IMF FA composition were analyzed. Animals with a higher proportion of MUFA and lower percentage of PUFA, exhibited a lower expression of β-oxidation genes in liver, an increased expression of lipogenic genes in adipose tissue and a metabolic shift towards lipogenesis and glucose uptake in muscle [61].

Finally, *FABP4g.2634_2635insC* polymorphism is located in a target binding site for PPARG, a well-known regulator of adipogenesis and FA metabolism. In a previous study of our group we identified the *PPARG* gene as a major regulator for fatness and growth traits in the same population [62]. PPARG has largely been reported as an

essential transcription factor for adipogenesis that modules *FABP4* gene expression [63,64,65]. Furthermore, *PPARG* is known to be involved in lipid metabolism and it is selectively expressed in pig adipose tissue. In mice, *PPARG*-knockdown (isoforms 1 and 2) inhibits adipogenesis and *FABP4* gene expression [66]. In addition, other studies have reported evidences for PPARG regulation via enhancing an intron 1 binding site [67]. Here, we hypothesize that PPARG may regulate *FABP4* expression via enhancing an intron 1 binding site. However, we cannot discard other polymorphisms on this gene as causative for the SSC4 QTL and further functional studies are required.

Conclusions

Association and QTL scan results showed that *FABP4:g.2634_2635insC* was the most significant marker for C16:1(n-7)), C20:3(n-6) FA, and PUFA composition in *Longissimus dorsi* muscle. In addition, *FABP4:g.2634_2635insC* polymorphism was associated with *FABP4* gene expression in BF, but not in muscle. These results suggest the *FABP4:g.2634_2635insC* polymorphism as being a strong candidate polymorphism for FA composition in muscle. The *FABP5:g.3000T>G* SNP was the most significant marker for IMF percentages of C18:1(n-9), C18:2(n-6), and MUFA. However, this variant was not associated with *FABP5* gene expression, being *FABP5* gene expression in muscle regulated by other genomic regions located on SSC4, SSC6, SSC9 and SSC13.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JMF, MB and AIF conceived and designed the experiment; JMF was the principal investigator of the project. AIF and JMF collected samples. AC, AP and MB performed the DNA and RNA isolation. AC, AP and MG performed the expression analysis. The polymorphisms genotyping was done by AC. AP and JMF performed the association analysis. AP and MB did the functional analysis of eGWAS regions. AP, MB and JMF wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was funded by MINECO AGL2011-29821-C02 and AGL2014-56369-C2 grants. A. Puig-Oliveras was funded by a "Personal Investigador en Formació" (PIF) grant from Universitat Autònoma de Barcelona (458-01-1/2011). The authors gratefully acknowledge J.L. Noguera and other researchers of IRTA (Institut de Recerca i Tecnologia Agroalimentàries) for animal material contribution.

References

1. Serra X, Gil F, Pérez-Enciso M, Oliver M, Vázquez J, Gispert M, *et al.* A comparison of carcass meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Lives Prod Sci 1998, 56:215–223.
2. Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, *et al.* Fat deposition fatty acid composition and meat quality: A review. Meat Sci 2008, 78:343-358.
3. Zhang W, Xiao S, Samaraweera H, Lee EJ, Ahn DU. Improving functional value of meat products. Meat Sci 2010, 66(1):14-31.
4. Maharani D, Park HB, Lee JB, Yoo CK, Lim HT, Han SH, *et al.* Association of the gene encoding stearoyl-CoA desaturase (SCD) with fatty acid composition in an intercross population between Landrace and Korean native pigs. Mol Biol Rep 2013, 40:73-80.
5. Wood JD, Nute GR, Richardson RI, Whittington FM, Southwood O, Plastow R, *et al.* Effects of breed diet and muscle on fat deposition and eating quality in pigs. Meat Sci 2004, 67:651-667.
6. Pérez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, Folch JM, *et al.* A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross. J Anim Sci 2000, 78:2525–2531.
7. Clop A, Ovilo C, Pérez-Enciso M, Cercos A, Tomas A, Fernandez A, Coll A, Folch JM, Barragan C, Diaz I, *et al*. Detection of QTL affecting fatty acid composition in the pig. Mamm Genome 2003, 14:650–656.

8. Mercadé A, Estellé J, Noguera JL, Folch JM, Varona L, Silió L, *et al.* On growth fatness and form: A further look at porcine Chromosome 4 in an Iberian × Landrace cross. Mamm Genome 2005, 16:374–382.
9. Fernández AI, Pérez-Montarelo D, Barragán C, Ramayo-Caldas Yl, Ibáñez-Escriche N, Castelló A, *et al.* Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 Beadchip. BMC Genetics 2012, 13:41.
10. Ramayo-Caldas Y, Mercadé A, Castelló A, Yang B, Rodríguez C, Alves E, *et al.* Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. J Anim Sci 2012, 90:2883–2893.
11. Muñoz M, Rodríguez MC, Alves E, Folch JM, Ibañez-Escriche N, Silió L, et al. Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. BMC genomics 2013, 14:845.
12. Kim Y, Kong M, Nam YJ, Lee C. A Quantitative Trait Locus for Oleic Fatty Acid Content on Sus scrofa Chromosome 7. J Hered 2006, 97:535–537.
13. Nii M, Hayashi T, Tani F, Niki A, Mori N, Fujishima-Kanaya N, *et al.* Quantitative trait loci mapping for fatty acid composition traits in perirenal and back fat using a Japanese wild boar × Large White intercross. Anim Genet 2006, 37:342–347.
14. Guo T, Ren J, Yang K, Ma J, Zhang Z, Huang L. Quantitative trait loci for fatty acid composition in longissimus dorsi and abdominal fat: results from a White Duroc × Erhualian intercross F2 population. Anim Genet 2009, 40 185–191.
15. Gerbens F, Jansen A, van Erp AJM, Harders F, Meuwissen TH, Rettenberger G*, et al.* The adipocyte fatty acid-binding protein locus: characterization and association with intramuscular fat content in pigs. Mamm Genome 1998, 9:1022-1026.
16. Gerbens F, de Koning DJ, Harders FL, Meuwissen TH, Janss LL, Groenen MA, *et al.* The effect of adipocyte and heart fatty acid-binding protein genes on intramuscular fat and backfat content in Meishan crossbred pigs. J Anim Sci 2000, 78:552–559.
17. Gerbens F, Verburg FJ, Van Moerkerk HT, Engel B, Buist W, Veerkamp JH, *et al.* Associations of heart and adipocyte fatty acid-binding protein gene expression with intramuscular fat content in pigs. J Anim Sci 2001, 79:347–354.

18. Estellé J, Pérez-Enciso M, Mercadé A, Varona L, Alves E, Sánchez A, *et al.* Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross. Anim Genet 2006, 37:589–591.
19. Mercadé A, Pérez-Enciso M, Varona L, Alves E, Noguera JL, Sánchez A, *et al.* Adipocyte fatty-acid binding protein is closely associated to the porcine FAT1 locus on chromosome 4. J Anim Sci 2006, 84:2907–2913.
20. Gao Y, Zhang YH, Zhang S, Li F, Wang S, Dai L, *et al.* Association of A-FABP gene polymorphism in intron 1 with meat quality traits in Junmu No. 1 white swine. Gene 2011, 487:170-173.
21. Atshaves BP, Martin GG, Hostetler HA, McIntosh AL, Kier AB, Schroeder F. Liver fatty acid-binding protein and obesity. J Nutr Biochem 2010, 21:1015–1032.
22. Fischer H, Gustafsson T, Sundberg CJ, Norrbom J, Ekman M, Johansson O, et al. Fatty acid binding protein 4 in human skeletal muscle. Biochem Biophys Res Commun 2006, 346:125-130.
23. Lee EJ, Lee HJ, Kamli MR, Pokharel S, Bhat AR, Lee YH, *et al.* Depot-specific gene expression profiles during differentiation and transdifferentiation of bovine muscle satellite cells and differentiation of preadipocytes. Genomics 2012, 100:195–202.
24. Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, *et al.* Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. Plos One 2009, 4:e6524.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 2007, 81:559–575
26. NAGRP Community Data Repositoryfor Livestock Animal Genomics. Groenen M. In: Illumina Pig 60K SNPs mapped to pig genome assembly 10.2. http://www.animalgenome.org/repository/pig/. Accessed Jan 2015.
27. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 2005, 21(13):2933-42.

28. Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acid Res. 1998, 16(22):10881-90.
29. Muñoz M, Alves E, Ramayo-Caldas Y, Casellas J, Rodríguez C, Folch JM, *et al.* Recombination rates across porcine autosomes inferred from high-density linkage maps. Anim Genet 2012, 43(5):620-3.
30. Ron M, Lwein H, Da Y, Band M, Yanai A, Blank Y, *et al.* Prediction of informativeness for microsatellite markers among progeny of sires used for detection of economic trait loci in dairy cattle. Anim Genet 1995, 26:439-441.
31. Green P, Falls K, Crooks S. Documentation for CRI-MAP version 2.4. St. Louis MO: Washington School of Medicine 1990.
32. Pérez-Enciso M, Misztal I. Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 2011, 12:202.
33. Ihaka R, Gentleman R. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 1996, 5(3):299-314.
34. Cunningham F, Amode MR, Beal K, Konstantinos B, Brent S, Carvalho-Silva D, *et al.* Ensembl 2015. Nucleic Acids Research 2015 43:D662-669. http://www.ensembl.org. Accessed 20 May 2015.
35. Øvergård AC, Nerland A, Patel S. Evaluation of potential reference genes for real time RT-PCR studies in Atlantic halibut (*Hippoglossus Hippoglossus L.*); during development in tissues of healthy and NNV-injected fish and in anterior kidney leucocytes. BMC Mol Biol 2010, 11:36.
36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001, 25(4):402-8.
37. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci 2003, 100 9440–9445.
38. Moller M, Berg F, Riquet J, Pomp D, Archibald A, Anderson S, *et al.* High-resolution comparative mapping of pig Chromosome 4 emphasizing the FAT1 region. Mamm Genome 2004, 15:717–731.
39. Cánovas A, Quintanilla R, Amills M, Pena R. Muscle tanscriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics 2010, 11:372.

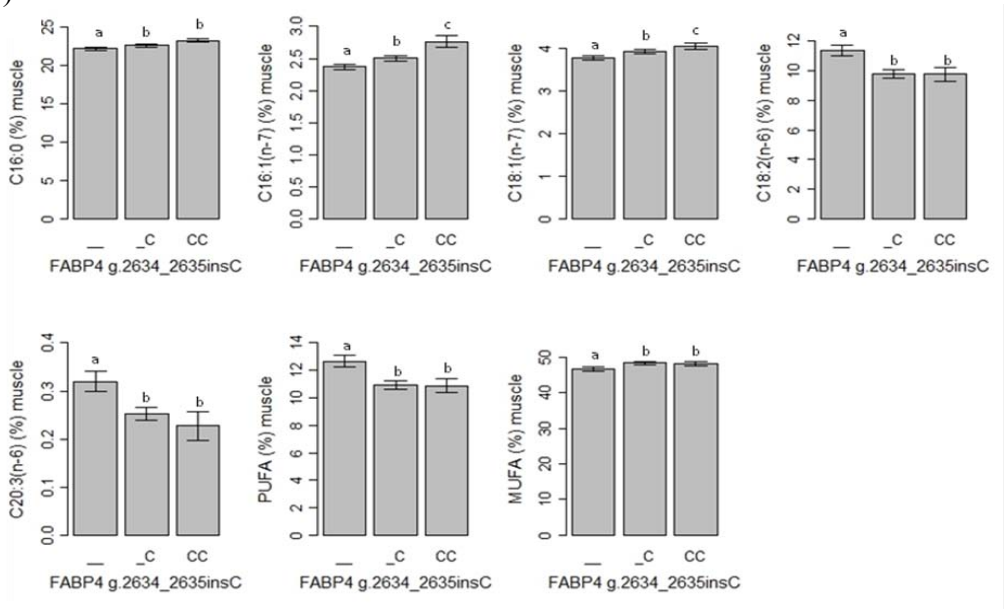
40. Pena R, Noguera JL, Casellas J, Díaz I, Fernández AI, Folch JM, *et al.* Transcriptional analysis of intramuscular fatty acid composition in the longissimus thoracis muscle of Iberian × Landrace back-crossed pigs. Animal Genetics 2013, 44:648-660.
41. Coe NR, Bernlohr DA. Physiological properties and functions of intracellular fatty acid-binding proteins. Biochim Biophys Acta 1998, 1391:287-306.
42. Queipo-Ortuño MI, Escoté X, Ceperuelo-Mallafré V, Garrido-Sanchez L, Miranda M, Clemente-Postigo M, *et al.* FABP4 dynamics in obesity: discrepancies in adipose tissue and liver expression regarding circulating plasma levels. Plos One 2012, 7(11):e48605.
43. Takeda Y, Kang HS, Lih FB, Jiang H, Blaner WS, Jetten A. Retinoid acid-related orphan receptor γ, RORγ, participates in diurnal transcriptional regulation of lipid metabolic genes. Nucleic Acids Research 2014, 42(16): 10448–10459.
44. Tseng YH, Butte AJ, Kokkotou E, Yechoor VK, Taniguchi CM, Kriauciunas KM, *et al.* Prediction of preadipocyte differentiation by gene expression reveals role of insulin receptor substrates and necdin. Nat Cell Biol 2005, 7(6):601-11.
45. Eguchi J, Wada J, Hida K, Zhang H, Matsouka T, Baba M, *et al.* Identification of adipocyte adhesion molecule (ACAM), a novel CTX gene family, implicated in adipocyte maturation and development of obesity. Biochemical Journal 2005, 387:343-353.
46. Schmitz G, Ecker J. The opposing effects of n-3 and n-6 fatty acids. Prog Lipid Res 2008, 47(2):147-55.
47. Furuhashi M, Hotamisligil GS. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. Nat Rev Drug Discov 2008, 7(6):489-503.
48. Kralisch S, Ebert T, Lossner U, Jessnitzer B, Stumvoll M, Fasshauer M. Adipocyte fatty acid-binding protein is released from adipocytes by a non-conventional mechanism. Int J Obes 2014, 38(9):1251-4.
49. Thumser AE, Moore JB, Plant NJ. Fatty acid binding proteins: tissue-specific functions in health and disease. Curr Opin Clin Nutr Metab Care 2014, 17(2):124-9.
50. Doran O, Moule SK, Teye GA, Whittington FM, Hallet KG, Wood JD. A reduced protein diet induces stearoyl-CoA desaturase protein expression in pig muscle but not in subcutaneous adipose tissue: relationship with intramuscular lipid formation. British Journal of Nutrition 2006, 95(3):609-617.

51. Ntambi JM, Miyazaki M, Stoehr JP, Kendziorski CM, Yandell BS, Song Y, *et al.* Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. PNAS 2002, 99(17):11483.
52. Cao H, Gerhold K, Mayers JR, Wiest MM, Watkins SM, Hotamisligil GS. Identification of a Lipokine a Lipid Hormone Linking Adipose Tissue to Systemic Metabolism. Cell 2008, 134:933–944.
53. Khalyfa A, Bhushan B, Hegazi M, Kim J, Kheirandish-Gozal L, Bhattacharjee R, *et al.* Fatty acid binding protein 4 gene variants and childhood obesity: potential implications for insulin sensitivity and CRP levels. Lipids in Health and Disease 2010, 9:18.
54. Bag S, Ramaiah S, Anbarasu A. Fabp4 is central to eight obesity associated genes: a functional gene network-based polymorphic study. J Theor Biol 2015, 7(364):344-54.
55. Hotamisligil GS, Johnson RS, Distel RJ, Ellis R, Papaioannou VE, Spiegelman BM. Uncoupling of Obesity from Insulin Resistance Through a Targeted Mutation in aP2 the Adipocyte Fatty Acid Binding Protein. Science 1996, 274:1377–1379.
56. Uysal KT, Scheja L, Wiesbrock SM, Bonner-Weir S, Hotamisligil GS. Improved glucose and lipid metabolism in genetically obese mice lacking aP2. Endocrinology 2000, 141:3388-3396.
57. Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Estellé J, Castelló A, Alves E, *et al.* Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics 2013, 14:843.
58. Syamsunarno MRAA, Iso T, Hanaoka H, Yamaguchi A, Obokata M, Koitabashi N*, et al.* A critical role of Fatty Acid Binding Protein 4 and 5 (FABP4/5) in the systemic response to fasting. Plos One 2013, 8(11):e79386.
59. Maeda K, Cao H, Kono K, Gorgun CZ, Furuhashi M, Uysal KT, *et al.* Adipocyte/macrophage fatty acid binding proteins control integrated metabolic responses in obesity and diabetes. Cell Metab 2005, 1:107–119.
60. Ramayo-Caldas Y, Mach N, Esteve-Codina A, Corominas J, Castelló A, Ballester M, *et al.* Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics 2012, 13:547.

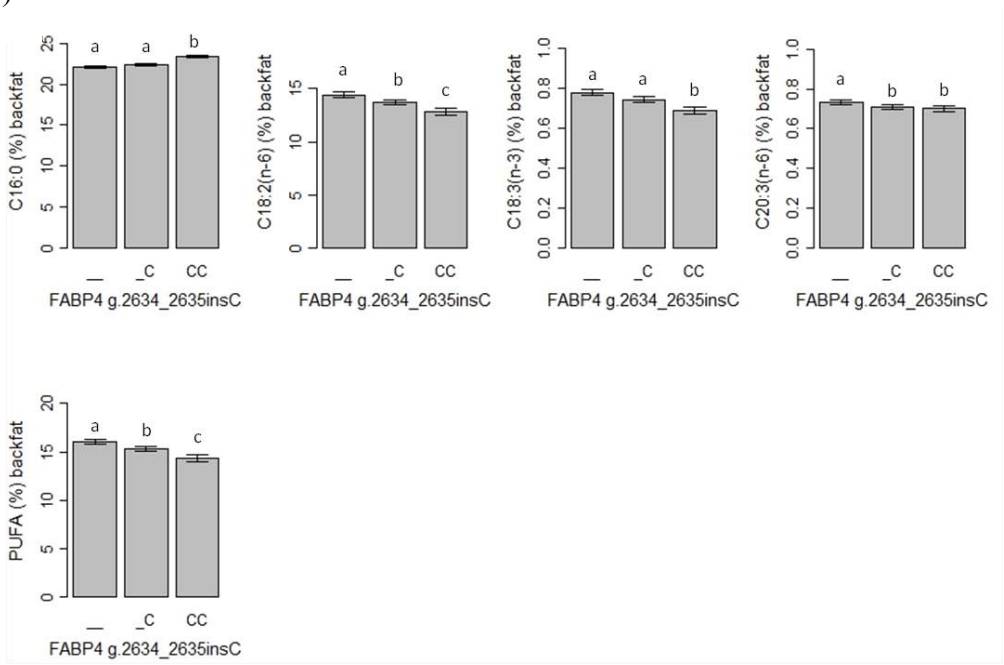
61. Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, *et al.* Differences in muscle transcriptome among pigs phenotipically extreme for fatty acid composition. Plos One 2014a, 9(7):e103668.
62. Puig-Oliveras A, Ballester M, Corominas J, Revilla M, Estellé J, Fernández AI, *et al.* A co-association network analysis of the genetic determination of pig conformation growth and fatness. Plos One 2014b, 9(12):e114862.
63. Grontved L, Mandrup S. PPARG: Peroxisome proliferator-activated receptor gamma. In: The Transcription Factor Encyclopedia. Genome Biology 2012, 13:R24.http://www.cisreg.ca/tfe/. Accessed 25 June 2015.
64. Lehrke M, Lazar MA. The many faces of PPARgamma. Cell 2005, 123(6):993-9.
65. Shin J, Li B, Davis ME, Suh Y, Lee K. Comparative analysis of fatty acid binding protein 4 promoters: Conservation of PPAR binding sites. J Anim Sci 2009, 87(12):3923-34.
66. Strand DW, Jiang M, Murphy TA, Konvinse KC, Franco OE, Wang Y, et al. PPARγ isoforms differentially regulate metabolic networks to mediate mouse prostatic epithelial differentiation. Cell Death and Disease 2012, 3:e361.
67. Bugge A, Siersbæk M, Madsen MS, Göndör A, Rougier C, Mandrup S. A novel intronic Peroxisome Proliferator-activated Receptor γ enhancer in the uncoupling protein (UCP) 3 gene as a regulator of both UCP2 and -3 expression in adipocytes. J Biol Chem 2010, 285(23):17310-7.

Figures

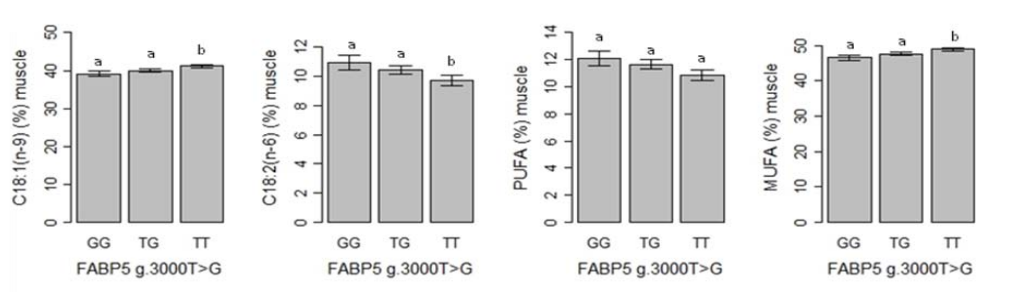
A)


C16:0 (%) muscle
C16:1(n-7) (%) muscle
C18:1(n-7) (%) muscle
C18:2(n-6) (%) muscle
C20:3(n-6) (%) muscle
PUFA (%) muscle
MUFA (%) muscle
__
_C
CC
FABP4 g.2634_2635insC

B)


C16:0 (%) backfat
C18:2(n-6) (%) backfat
C18:3(n-3) (%) backfat
C20:3(n-6) (%) backfat
PUFA (%) backfat
__
_C
CC
FABP4 g.2634_2635insC

C)


C18:1(n-9) (%) muscle
C18:2(n-6) (%) muscle
PUFA (%) muscle
MUFA (%) muscle
GG
TG
TT
FABP5 g.3000T>G

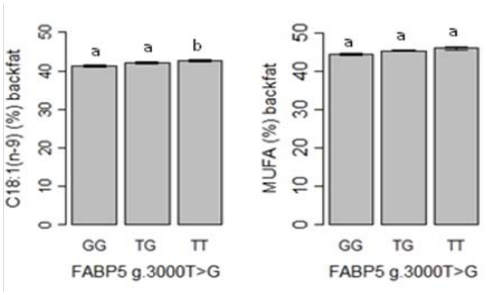


Figure 1. FA composition in IMF and BF classified according to genotypes of *FABP4*:*g.2634_2635insC* and *FABP5*:*g.3000T>G* polymorphisms. (A) FA content in muscle and *FABP4*:*g.2634_2635insC* genotypes, (B) FA content in BF and *FABP4*:*g.2634_2635insC* genotypes, (C) FA content in muscle and *FABP5*:*g.3000T>G* genotypes, and (D) FA content in BF and *FABP5*:*g.3000T>G* genotypes. Data represent means ± standard error of mean (SEM). Values with different superscript letter (a, b and c) indicate significant differences between groups (p-value < 0.05) as determined by a linear model in R.

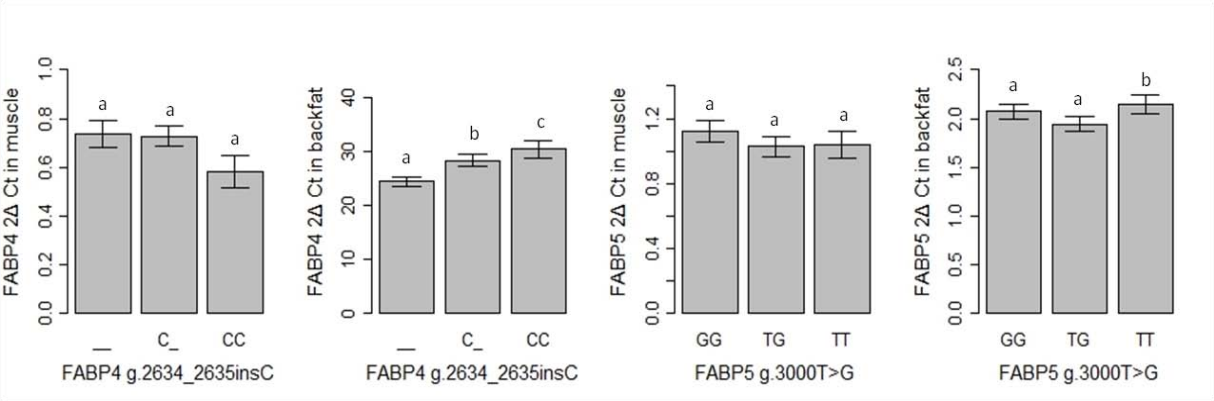
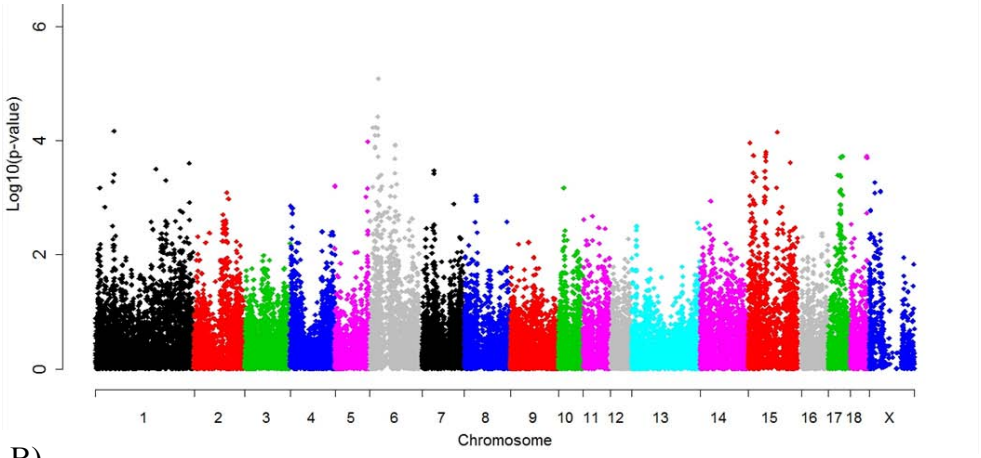
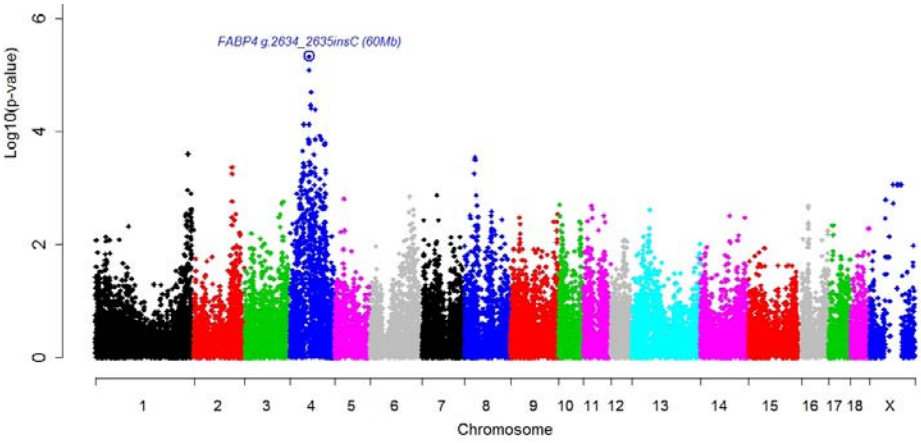


Figure 2. *FABP4* and *FABP5* gene expression in muscle and backfat classified according the *FABP4*:*g.2634_2635insC* and *FABP5*:*g.3000T>G* genotypes. Data represent means ± standard error of mean (SEM). Values with different superscript letter (a, b and c) indicate significant differences between groups (p-value < 0.05) as determined by a linear model in R.

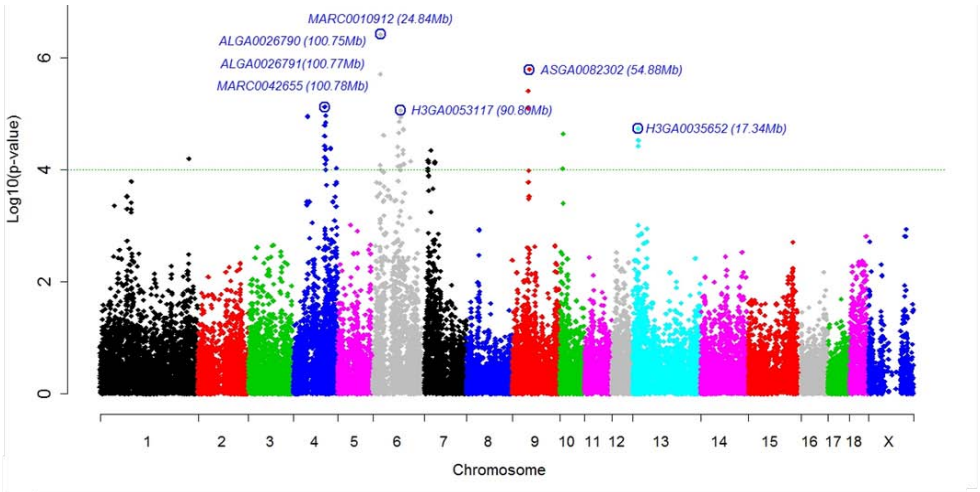
A)


Log10(p-value)
0
2
4
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

B)


FABP4 g.2634_2635insC (60Mb)
Log10(p-value)
0
2
4
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

C)


MARC0010912 (24.84Mb)
ALGA0026790 (100.75Mb)
ALGA0026791(100.77Mb)
MARC0042655 (100.78Mb)
ASGA0082302 (54.88Mb)
H3GA0053117 (90.88Mb)
H3GA0035652 (17.34Mb)
Log10(p-value)
0
2
4
6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

D)

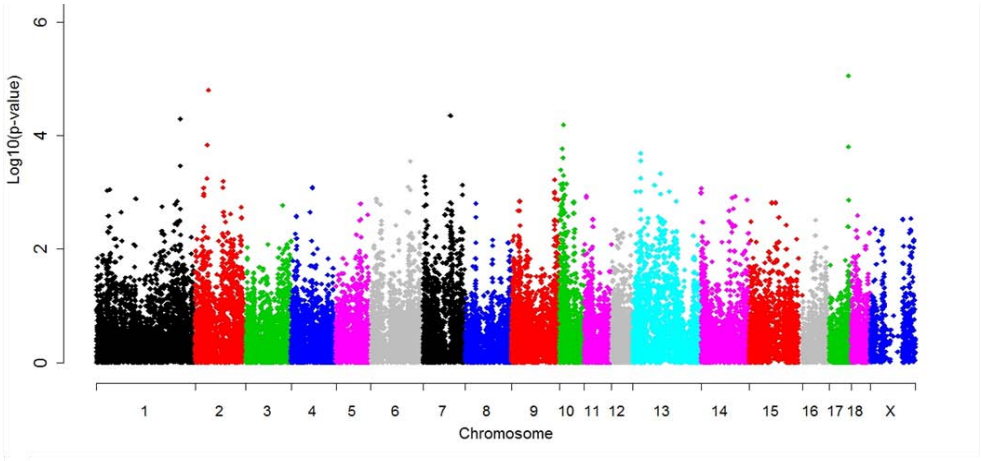


Figure 3. Plot of *FABP4* and *FABP5* gene expression GWAS in BF and muscle tissues. Positions in Mb are relative to *Sscrofa10.2* assembly of the pig genome. Plot of eGWAS for (A) *FABP4* gene expression in muscle (B) *FABP4* gene expression in BF (C) *FABP5* gene expression in muscle and (D) *FABP5* gene expression in BF. Horizontal dashed line indicates the genome-wide significance level (FDR-based q-value ≤ 0.05).

Tables

Table 1. Association of *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms with FA composition traits in IMF and backfat tissues.

Fatty Acid Composition (%)	Polymorphism association				Mean (SD)
	FABP4 *g.2634_2635insC*	Additive genetic effect	*FABP5* *g.3000T>G*	Additive genetic effect	
Intramuscular					
C16:0	1.77×10^{-3} *	0.418	8.56×10^{-1}	0.418	22.568 (1.205)
C16:1(n-7)	1.29×10^{-6} ***	0.195	3.55×10^{-2}	0.102	2.498 (0.389)
C18:1(n-7)	1.00×10^{-3} *	0.144	1.34×10^{-2}	0.118	3.883 (0.359)
C18:1(n-9)	4.47×10^{-2}	0.685	7.61×10^{-4} **	1.223	40.065 (2.892)
C18:2(n-6)	1.25×10^{-3} *	-0.929	9.71×10^{-4} **	-1.020	10.394 (2.482)
C20:3(n-6)	1.21×10^{-4} **	-0.046	2.36×10^{-2}	-0.040	0.274 (0.131)
MUFA	8.22×10^{-3} *	0.981	3.24×10^{-5} ***	1.469	47.951 (3.193)
PUFA	1.58×10^{-3} *	-1.290	2.48×10^{-3} *	-1.333	13.405 (3.444)
Backfat					
C16:0	4.22×10^{-4} **	0.454	6.86×10^{-1}	0.000	22.463 (1.163)
C18:1(n-9)	2.89×10^{-1}	0.205	5.59×10^{-3} *	0.564	41.997 (1.839)
C18:2(n-6)	4.65×10^{-4} **	-0.674	4.93×10^{-2}	-0.409	13.853 (1.762)
C18:3(n-3)	5.34×10^{-3} *	-0.031	3.59×10^{-1}	-0.011	0.745 (0.106)
C20:3(n-6)	7.99×10^{-3} *	-0.023	1.58×10^{-1}	-0.013	0.718 (0.080)
MUFA	2.70×10^{-1}	0.227	9.19×10^{-3} *	0.552	45.393 (1.943)
PUFA	3.98×10^{-4} **	-0.738	5.70×10^{-2}	-0.430	15.429 (1.891)

*p-values<0.01 ; **p-values<0.001 ; ***p-values<0.0001

Table 2. QTL analysis and likelihood ratio tests comparing the QTL and gene polymorphism effect (*FABP4-FABP5* genes).

Fatty Acid Composition (%)	QTL[a]		QTL + Gene polymorphism[b]		
	Pos (cM)[c]	P_{QTL}	P_{QTL}[b]	*FABP4* *g.2634_2635insC*	*FABP5* *g.3000T>G*
Intramuscular fat					
C16:1(n-7)	45-54	5.91×10^{-5} ***	8.31×10^{-4} ***	1.65×10^{-6} ***	
C18:1(n-9)	50	3.34×10^{-4} ***	3.34×10^{-1}		7.55×10^{-4} ***
C18:2(n-6)	34-71	3.25×10^{-6} ***	3.23×10^{-3} **		1.32×10^{-4} ***
C20:3(n-6)	50	3.46×10^{-4} ***	1.16×10^{-1}	1.11×10^{-4} ***	
PUFA	37-53	4.64×10^{-6} ***	5.08×10^{-2}	1.62×10^{-3} **	
MUFA	36-57	3.88×10^{-5} ***	1.89×10^{-2} *		3.52×10^{-5} ***
Backfat					
C18:2(n-6)	48-50	2.11×10^{-4} ***	2.84×10^{-4} ***	5.59×10^{-4} ***	
PUFA	48-50	2.05×10^{-4} ***	1.96×10^{-2} *	5.06×10^{-4} ***	

[a] A classical QTL model (1)

[b] A model including both QTL and gene polymorphism effects. For each trait, only the most significant of the *FABP4:g.2634_2635insC* and *FABP5:g.3000T>G* polymorphisms was included in the model. Likelihood ratio tests for QTL and SNP effects were calculated comparing the appropriate full and reduced models.

[c] QTL position in cM; *FABP4* and *FABP5* genes are located at 50 cM in the linkage map.

* p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.

Additional files

Additional file 1: Table S1. Markers in SSC4 for QTL scan analysis and their position in the linkage map (cM) and the physical map (bp).

Additional file 2: Table S2. Primers used for *FABP4* and *FABP5* gene expression quantification by RT-qPCR.

Additional file 3: Figure S1. QTL scan with significant QTL regions. QTL scan for C16:1(n-7), C18:1(n-9), C18:2(n-6) and C20:3(n-6) FAs, PUFA and MUFA IMF content and C18:2(n-6) FA and PUFA backfat content.

Additional file 4: Figure S2. Multi-alignment of porcine, cattle, and human *FABP4* gene sequence of exon 1 and proximal intron 1. This region harbours *FABP4:g.2634_2635insC* polymorphism and PPARG putative transcription factor binding site affected by this polymorphism. The grey dotted box indicates *FABP4* exon 1 region, the black dotted box indicates the PPARG putative transcription binding site, and the black continuous box *FABP4:g.2634_2635insC* polymorphism in intron 1.

Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition

Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, Casellas J, Folch JM, Ballester M

PLoS One. 2014; 9(7): e103668. doi: 10.1371/journal.pone.0099720

OPEN ACCESS Freely available online

PLOS ONE

Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition

Anna Puig-Oliveras[1,2]*, Yuliaxis Ramayo-Caldas[3], Jordi Corominas[1,2], Jordi Estellé[3,4,5], Dafne Pérez-Montarelo[6], Nicholas J. Hudson[7], Joaquim Casellas[2,8], Josep M. Folch[1,2], Maria Ballester[1,2]

1 Departament de Genètica Animal, Centre de Recerca en Agrigenòmica (CRAG), Bellaterra, Spain, **2** Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain, **3** Génétique Animale et Biologie Intégrative UMR1313 (GABI), Institut National de la Recherche Agronomique (INRA), Jouy-en-Josas, France, **4** Génétique Animale et Biologie Intégrative UMR1313 (GABI), AgroParisTech, Jouy-en-Josas, France, **5** Laboratoire de Radiobiologie et Etude du Génome (LREG), Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Jouy-en-Josas, France, **6** Departamento de Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain, **7** Computational and Systems Biology, Commonwealth Scientific and Industrial Research Organisation (CSIRO) Animal, Food and Health SciencesQLD, Brisbane, Australia, **8** Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologies Agroalimentàries (IRTA), Lleida, Spain

Abstract

Background: Besides having an impact on human health, the porcine muscle fatty acid profile determines meat quality and taste. The RNA-Seq technologies allowed us to explore the pig muscle transcriptome with an unprecedented detail. The aim of this study was to identify differentially-expressed genes between two groups of 6 sows belonging to an Iberian × Landrace backcross with extreme phenotypes according to FA profile.

Results: We sequenced the muscle transcriptome acquiring 787.5 M of 75 bp paired-end reads. About 85.1% of reads were mapped to the reference genome. Of the total reads, 79.1% were located in exons, 6.0% in introns and 14.9% in intergenic regions, indicating expressed regions not annotated in the reference genome. We identified a 34.5% of the intergenic regions as interspersed repetitive regions. We predicted a total of 2,372 putative proteins. Pathway analysis with 131 differentially-expressed genes revealed that the most statistically-significant metabolic pathways were related with lipid metabolism. Moreover, 18 of the differentially-expressed genes were located in genomic regions associated with IMF composition in an independent GWAS study in the same genetic background. Thus, our results indicate that the lipid metabolism of FAs is differently modulated when the FA composition in muscle differs. For instance, a high content of PUFA may reduce FA and glucose uptake resulting in an inhibition of the lipogenesis. These results are consistent with previous studies of our group analysing the liver and the adipose tissue transcriptomes providing a view of each of the main organs involved in lipid metabolism.

Conclusions: The results obtained in the muscle transcriptome analysis increase the knowledge of the gene regulation of IMF deposition, FA profile and meat quality, in terms of taste and nutritional value. Besides, our results may be important in terms of human health.

Citation: Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, et al. (2014) Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition. PLoS ONE 9(6): e99720. doi:10.1371/journal.pone.0099720

Editor: Hervé Guillou, INRA, France

Received December 23, 2013; **Accepted** May 19, 2014; **Published** June 13, 2014

Funding: This work has been funded by MICINN project AGL2011-29821-C02 (Ministerio de Economía y Competitividad) and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036). A. Puig-Oliveras was funded by a Personal Investigador en Formación (PIF) PhD grant from the Universitat Autònoma de Barcelona (458-01-1/2011). J. Corominas was funded by a Formación de Personal Investigador (FPI) PhD grant from Spanish Ministerio de Educación (BES-2009-018223), Y. Ramayo by a Formación del Profesorado Universitario (FPU) PhD grant (AP2008-01450). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: anna.puig@cragenomica.es

Introduction

High-throughput sequencing technologies are rapidly evolving and its application to transcriptome analysis (RNA-Seq), with the adapted bioinformatic tools, allow the exploration of the transcriptome in an unprecedented manner in terms of accuracy and data insight [1]. In addition, RNA-Seq technology is useful, not only to detect variation in gene expression patterns, but also to identify new isoforms, splicing events, and different promoter and polyadenylation signal usage. Currently, only a few RNA-Seq studies have been conducted in livestock species such as pigs [2–6].

According to the Food and Agriculture Organization (FAO) [7], pork is the major source of meat intake by human, accounting for the 43% of the consumed meat worldwide. The taste and the quality of the cooked and the cured meat products depend on the oxidative stability of the muscle which is related to the fatty acid (FA) composition [8,9]. Furthermore, it is well known that genetic and environmental factors such as diet, are responsible for FA composition variation [10]. Besides its influence on meat taste, the FA composition in muscle has taken additional importance due to their nutritional value and human health-related benefits [11,12], particularly for its effects on human diseases such as cancers,

coronary heart diseases and atherosclerosis [11]. It has been reported that omega-3 FAs, such as α-linolenic acid (C18:3 n-3), are associated with the reduction of low density lipoprotein (LDL) cholesterol and blood triacylglycerols, as well as with the modulation of immune functions and inflammatory processes [13,14]. Artificial selection to increase meat production in pigs has caused a reduction of intramuscular fat (IMF) and changes in meat FA composition in some breeds. Hereby, there is an increasing interest in the pork industry on producing meat products with a higher IMF content and with a healthier FA profile, while maintaining a reduced amount of backfat [15].

In a recent genome-wide association study (GWAS) [16], genomic regions associated with the IMF (*Longissimus dorsi*) FA composition were identified in a backcross population (BC1_LD; 25% Iberian and 75% Landrace). A combined linkage QTL scan and GWAS performed in the same backcross revealed significant pleiotropic regions with effects on both IMF and backfat tissues [17,18]. Moreover, the transcriptome of the other two major organs regulating lipid metabolism, liver and adipose tissue (backfat), have been studied using RNA-Seq in the BC1_LD animals [4,5]. In these studies, a shift towards the oxidation of FAs in liver [4] and an inhibition of *de novo* lipogenesis in adipose tissue [5] was observed in animals with higher content of polyunsaturated FA (PUFA). Since the adipose and liver tissues have previously been analysed using animals belonging to the same population, with the addition of the muscle transcriptome we aim to have a more complete view of the genetic regulation of lipid metabolism in pigs [4,5]. The goal of the current study was to identify differentially-expressed genes and pathways in the *Longissimus dorsi* muscle of Iberian × Landrace backcrossed pigs with extreme phenotypes for muscle FA profile to better understand the differences in this meat quality trait.

Results

Phenotypic differences among the analysed animals

In a previous work [4], a Principal Component Analysis (PCA) was performed to select animals of an Iberian × Landrace backcross (BC1_LD) with extreme phenotypes for IMF FA composition. Using the same classification and the first principal component, six females belonging to the extreme High (H) group and six from the Low (L) group were selected for muscle RNA-Seq analysis. Animal selection considered the parental genetic diversity according to the pedigree. Significant statistical differences (*P*-value <0.05) were identified between the H and L groups in 18 out of 26 evaluated traits (Table 1). The H group had, in comparison to the L group, a higher content of PUFA including linolenic (C18:2 n-6), α-linolenic, eicosadienoic (C20:2 n-6), eicosatrienoic (C20:3 n-6) and arachidonic (C20:4 n-6) FAs. Conversely, the L group had a higher content of monounsaturated FA (MUFA) like palmitoleic (C16:1 n-7) and oleic (C18:1 n-9) FAs and saturated FAs (SFA) such as myristic (C14:0) and palmitic (C16:0) FAs. The two groups of pigs did not differ significantly in either IMF content or backfat thickness.

Transcriptome analysis of swine muscle tissue

As described above, the *Longissimus dorsi* (LD) muscle transcriptome was sequenced in twelve sows (H = 6, L = 6) with extreme phenotypes for intramuscular FA composition. A total amount of 787.5 M of 75 bp paired-end reads were acquired from the RNA-Seq experiment. Sequence alignment was performed against the reference pig genome (Sscrofa10.2) by using Tophat [19]. About 85.1% (76.5%–86.6%) of reads were mapped to the reference genome, of which 14.5% (12.4%–16.1%) did not map to unique genomic locations. A total of 85.1% (84.0%–87.6%) of the mapped reads correspond to annotated genes, 79.1% (77.5%–84.1%) of them were located in exons and 6.0% (3.6%–6.8%) in introns. The remaining 14.9% (12.4%–16.0%) of reads mapped to intergenic regions, indicating that they were not annotated in the reference genome (Table S1).

The transcripts generated when assembling the short reads with Cufflinks [20] resulted in a mean of 43,255 transcripts expressed in muscle (Table S2). Transcripts were classified in different categories, being the most abundant the exonic transcripts (60.4%), the putative new isoforms (20.5%) and the intergenic transcripts (10.1%) (Table S2). A total of 9,887 new isoforms were identified corresponding to 9,805 known genes.

Transposable elements identification and novel coding gene discovery

The percentage of interspersed repeats identified with the Repeat Masker [21] in the intergenic transcripts was about 34.5%. Moreover, SINEs and LINEs were the most abundant repetitive elements identified (14.1% and 14.9%, respectively) (Table S3).

With the aim of improving the current porcine genome annotation, we took into account the intergenic transcripts identified with cufflinks (a mean of 4,440 transcripts) to determine whether these transcripts potentially codified for proteins. A total of 2,372 putative proteins were predicted by Augustus [22] corresponding to non-annotated transcripts of the Sscrofa10.2 genome assembly version. Among the 2,372 novel predicted proteins, only 1,406 (59.2%) had at least one orthologous gene identified with BLASTP option of Blast2GO, representing a total of 577 known genes (Table S4) [23]. These proteins corresponded to: 720 *Sus scrofa*, 17 *Homo sapiens* and 247 *Bos taurus in silico* predicted protein, and 476 *Sus scrofa*, 933 *Homo sapiens*, and 403 *Bos taurus* known proteins. The pig species was the only one showing a higher percentage of computationally predicted protein (60.2%) in comparison to known proteins (39.8%).

Moreover, 918 of the predicted novel proteins were successfully annotated with Blast2GO [23]. To summarize the functional annotation, a GO Slim analysis was performed. The most relevant molecular functions identified were "protein binding" (25%), "ion binding" (19%), "nucleic acid binding" (17%), "small molecular binding" (10%) and, interestingly, "lipid binding" (2%). These new transcripts were mainly involved in the following biological processes: "primary metabolic process" (9%), "cellular metabolic process" (9%), "macromolecule metabolic process" (8%) and "regulation of biological process" (7%). Using the Enzime code and KEGG, the main metabolic pathways represented were the "phosphatidylinositol signalling system" (12 sequences), "inositol phosphate metabolism" (11) and the "pyrimidine (9) and purine (8) metabolism".

Differential gene expression analysis

A high correlation ($r = 0.98$, *P*-value $<2.2\times10^{-16}$) between H and L groups in the mean gene expression was found, showing that most of the genes had a similar behaviour. A total of 11,945 transcripts were used to perform the differential expression analysis after filtering. Using EdgeR program [24], 314 genes were identified as significantly differentially expressed between H and L groups. Whereas, employing DESeq [25], 208 genes were detected. Figure 1 shows the *P*-value distribution and how among the transcripts accepted as differentially expressed the selected cut-off of *P*-value <0.01 is clearly departing from the expected *P*-value (equivalent to a FDR ≤ 0.12).

A total of 131 genes (Table S5) overlapping in both analyses were selected as differentially expressed between H and L groups

Table 1. Mean comparison between High and Low groups (six animals per group) of the traits included in the principal component analysis (PCA).

Characters	H Mean	L Mean	Significance	*P*-value
Carcass quality				
Carcass weight (CW) (kg)	66.22±10.52	70.50±7.91	NS	4.44×10^{-1}
Ham weight (HW) (kg)	18.79±2.32	18.97±2.41	NS	8.96×10^{-1}
Shoulder weight (SW) (kg)	6.6±0.95	6.32±0.97	NS	6.16×10^{-1}
Intramuscular fat (IMF) (%)	1.94±0.65	1.69±0.64	NS	5.25×10^{-1}
Fatty acids in intramuscular fat				
Saturated FA[a]				
Myristic acid (C14:0)	1.11±0.12	1.28±0.12	*	$3{,}08\times10^{-2}$
Palmitic acid (C16:0)	21.29±0.57	24.16±0.54	***	4.35×10^{-6}
Heptadecanoic acid (C17:0)	0.35±0.06	0.20±0.03	***	3.09×10^{-4}
Stearic acid (C18:0)	13.50±0.94	14.16±1.11	NS	2.91×10^{-1}
Arachidic acid (C20:0)	0.25±0.09	0.23±0.05	NS	5.83×10^{-1}
Monounsaturated FA[a]				
Palmitoleic acid (C16:1 n-7)	2.33±0.30	2.97±0.41	*	1.03×10^{-2}
Heptadecenoic acid (C17:1)	0.33±0.08	0.22±0.05	*	2.04×10^{-2}
Oleic acid (C18:1 n-9)	36.78±3.10	42.77±1.07	**	1.18×10^{-3}
Octadecenoic acid (C18:1 n-7)	3.85±0.20	4.14±0.27	NS	6.29×10^{-2}
Eicosenoic acid (C20:1 n-9)	0.82±0.13	0.82±0.07	NS	9.95×10^{-1}
Polyunsaturated FA[a]				
Linoleic acid (C18:2 n-6)	13.70±1.30	6.83±0.40	***	2.14×10^{-7}
α-Linolenic acid (C18:3 n-3)	1.14±0.42	0.47±0.07	**	3.23×10^{-3}
Eicosadienoic acid (C20:2 n-6)	0.61±0.16	0.38±0.05	**	8.39×10^{-3}
Eicosatrienoic acid (C20:3 n-6)	0.42±0.17	0.15±0.02	**	3.58×10^{-3}
Arachidonic acid (C20:4 n-6)	2.79±1.26	0.76±0.18	**	2.96×10^{-3}
Metabolic ratios				
Average Chain Length (ACL)	17.44±0.02	17.37±0.02	***	2.41×10^{-7}
Saturated FA (SFA)	36.49±1.08	40.02±1.28	***	4.20×10^{-4}
Monounsaturated FA (MUFA)	44.49±2.90	51.21±1.41	***	4.67×10^{-4}
Polyunsaturated FA (PUFA)	18.67±2.75	8.59±0.53	***	4.95×10^{-6}
Peroxidability index (PI)	30.92±6.66	13.53±1.02	***	8.60×10^{-5}
Double-bond index (DBI)	0.44±0.08	0.19±0.01	***	2.85×10^{-5}
Unsaturated index (UI)	0.89±0.06	0.70±0.01	***	1.82×10^{-5}

NS: *P*-value >0.05; * *P*-value <0.05; ** *P*-value <0.01; *** *P*-value <0.001
[a]The percentage of each FA, relative to the total FA
doi:10.1371/journal.pone.0099720.t001

and, thereafter, used for pathway analysis (Figure 2). Fifty genes had a higher expression and 81 a lower expression in the H group (in comparison with L group). Remarkably, eighteen (*CLCA4*, *ANGPT1*, *PLEKHH1*, *SDR16C5*, *PIK3R1*, *INTU*, *MAL2*, *NCEH1*, *PLN*, *C4orf29*, *FABP3*, *TBX3*, *MCT1*, *ESF1*, *POLR3GL*, *DBT*, *C6orf165* and *CHAC1*) of the 131 genes were also present in the annotated QTL intervals of a GWAS study for the IMF FA profile performed in the same population [16]. Three of these genes (*PIK3R1*, *NCEH1* and *FABP3*) have been directly related with lipid metabolism [26], being clear candidate genes to study the genetic contribution of IMF FA composition. Intriguingly, only two (*C6orf165* and *CHAC1*) of the 18 genes were over-expressed in the H group. Moreover, two of the differentially-expressed genes in muscle (*AQP7* and *FOS*) were also identified as differentially expressed in liver [4], and seventeen of them (*AQP4*, *SCD*, *PLEKHB1*, *CTSF*, *CIDEC*, *ALDOC*, *CXCL2*, *KIAA0408*, *SLPI*, *ALB*, *C14H10orf116*, *ITPR2*, *TRIP10*, *BANF1*, *HIF1AN*, *CHAC1* and *FHL3*) were identified as differentially expressed in adipose tissue [5]. In addition, three of the differentially-expressed genes in our analysis (*ATF3*, *ENAH* and *SLPI*) were also identified in a muscle microarray study of extreme animals for FA composition from the same backcross [27]. Other genes such as *DNAJA4*, *ANKRD1*, *MYH10* and *TNFRSF12A* were also common, but they were only detected by the DESeq program [25] in the RNA-Seq data.

Functional analysis

With the aim of having a more complete functional view of our differentially-expressed genes in the H and L groups, we used Babelomics [28] and Ingenuity Pathways Analysis [29] programs, who have related capabilities but use different databases. The top canonical pathways overrepresented according to IPA were related with nitric oxide signalling in the cardiovascular system

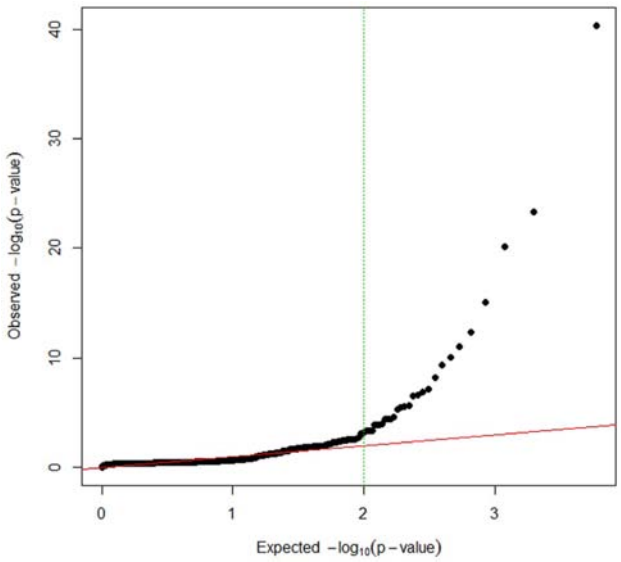


Figure 1. Q-Q plot representing the DESeq [25] ***P*-value distribution of the differentially expression analysis.** The expected distribution of the *P*-values is indicated with a red line, whereas black points represent the observed distribution. The selected cut-off is represented with a green discontinuous line (-log10 (*P*-value) >2).
doi:10.1371/journal.pone.0099720.g001

(7 genes, *P*-value $=7.75\times10^{-7}$) and endothelial nitric oxide synthase signalling (*eNOS*, 8 genes, *P*-value $=2.34\times10^{-6}$). On the other hand, using Babelomics we observed an overrepresentation of lipids and lipoproteins metabolism (6 genes, *P*-value $=1.64\times10^{-3}$), and also the peroxisome proliferator-activated receptors (*PPAR*, 4 genes, *P*-value $=7.25\times10^{-4}$) and the insulin (5 genes, *P*-value $=1.00\times10^{-3}$) signalling and the hemostasis (7 genes, *P*-value $=1.64\times10^{-3}$) pathway (Table 2).

Among the top molecular and cellular functions significantly overrepresented when comparing H relative to L groups with Babelomics, we identified the response to organic substance (18 genes, *P*-value $=3.8\times10^{-7}$), the muscle organ development (5 genes, *P*-value $=3.8\times10^{-5}$), the energy derivation by oxidation of organic compounds (5 genes, *P*-value $=1.0\times10^{-4}$) and the response to hormone stimulus (9 genes, *P*-value $=2.6\times10^{-4}$). Whereas with IPA, the most relevant functions were involved in lipid metabolism (30 genes, *P*-value $=1.04\times10^{-6}$), molecular transport (36 genes, *P*-value $=1.04\times10^{-06}$), small molecule biochemistry (47 genes, *P*-value $=1.04\times10^{-6}$), cell death and survival (38 genes, *P*-value $=1.55\times10^{-6}$), carbohydrate metabolism (30 genes, *P*-value $=2.25\times10^{-6}$), energy production (10 genes, *P*-value $=5.8\times10^{-5}$) and skeletal and muscular system development and function (23 genes, *P*-value $=2.47\times10^{-4}$) (Table S6).

Among the related specific functions for lipid metabolism, the top molecular functions identified with IPA were the oxidation (*ACADVL*, *ACOX2*, *FABP3*, *PLIN1*, *PLIN5*, *PON2*, *SCD*; *P*-value $=3.59\times10^{-4}$), accumulation (*ACADVL*, *AQP7*, *FH*, *IDH1*, *PLIN1*, *PON2*, *RETSAT*, *SCD*; *P*-value $=7.90\times10^{-04}$), synthesis

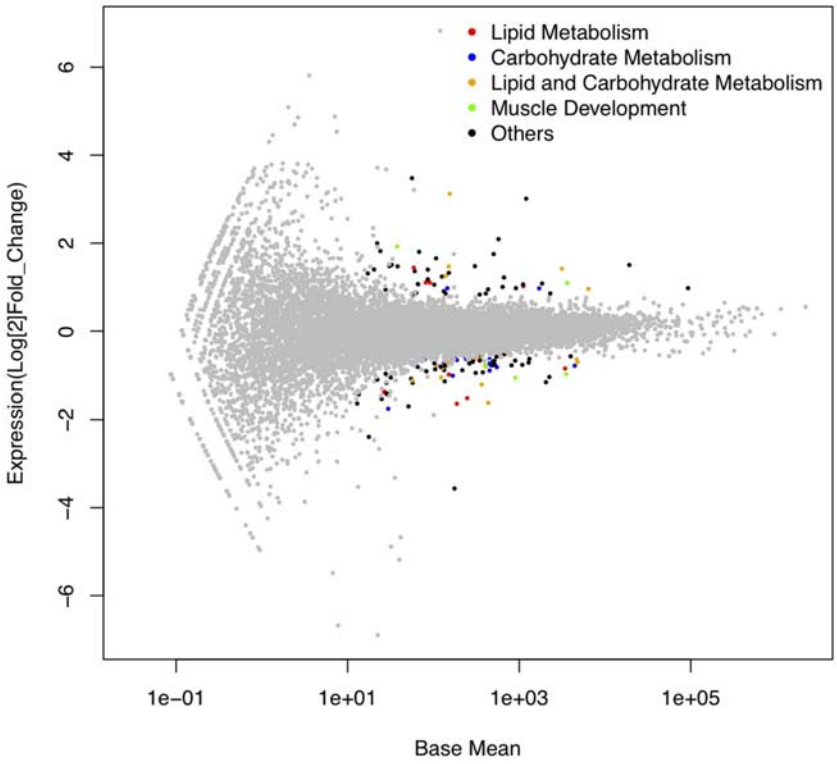


Figure 2. Plot of the 131 differentially-expressed genes identified between the two groups High and Low. X-axis values correspond to base mean expression values and y-axis values are the log2(fold change). The colour for the differentially-expressed genes is related to their function in lipid metabolism (red), carbohydrate metabolism (blue), both lipid and carbohydrate metabolism (orange), muscle development (green) or others (black).
doi:10.1371/journal.pone.0099720.g002

Table 2. Summary of the most significantly-overrepresented pathways of the differentially-expressed genes in muscle between High and Low groups for fatty acid composition traits.

Babelomics			IPA		
Category	**Genes**	***P*-value**	**Category**	**Genes**	***P*-value**
Metabolism of Lipids and Lipoproteins	*SCD, ACADVL, ACOX2, IDH1, IDI1, ALB*	5.12×10^{-4}	Nitric Oxid Signalling in the Cardiovascular System	*BDKRB2, PIK3R3, PRKG1, PLN, ITPR2, PIK3R1, ATP2A2*	7.75×10^{-7}
Alanine, Aspartate and Glutamate Metabolism	*ABAT, GOT1, ASNS*	6.38×10^{-4}	eNOS Signalling	*BDK4B2, PIK3R3, AQP7, PRKG1, ITPR2, PIK3R1, CHRNA9, AQP4*	2.34×10^{-6}
PPAR Signalling Pathway	*ACOX2, AQP7, FABP3, SCD*	7.25×10^{-4}	Clathrin-mediated Endocytosis Signalling	*PIK3R3, ALB, CD2AP, TF, PIK3R1, TFRC, ITGB6, FGF1*	2.95×10^{-5}
Insulin Signalling Pathway	*GYS2, PIK3R1, TRIP10, PIK3R3, PPP1R3C*	1.00×10^{-3}	*ILK* Signalling	*PIK3R3, FOS, RND3, PIK3R1, ITGB6, MYL6B, MYH7B, ACTN3*	2.95×10^{-5}
Hemostasis	*ANGPT1, PIK3R1, ALDOA, ITPR2, PLEK, TF, ALB*	1.64×10^{-3}	*CXCR4* Signalling	*PIK3R3, FOS, RND3, ITPR2, PIK3R1, MYL6B*	4.79×10^{-4}

doi:10.1371/journal.pone.0099720.t002

(*ACADVL, ACOX2, ALB, BDKRB2, CNTFR, FABP3, FGF1, FOS, IDI1, PIK3R1, PLIN1, PON2, SCD*; *P*-value $=2.28\times10^{-3}$), concentration (*DUSP1, EXTL1, FABP3, FOS, IDH1, NCEH1, PLIN1, PON2, PPP1R3C, SCD*; *P*-value $=2.48\times10^{-5}$) and homeostasis (*ACADVL, FABP3, GOT1, NCEH1, PIK3R1, PLIN1, SCD*; *P*-value $=1.96\times10^{-4}$) of lipids (Table S7). Other related pathways were identified such as concentration of bile acids (*ALB, ATF3, SCD*; *P*-value $=3.87\times10^{-4}$), obesity (*ABAT, AQP7 ARID5B, ATF3, DESP1, HBEGF, IDH1, PLIN1, RETSAT, SCD*; *P*-value $=5.76\times10^{-4}$), and insulin resistance (*ACOX2, ALB, AQP7, ATP2A2, FGF1, PIK3R1, PON2, PPP1R3C, SCD*; *P*-value $=5.41\times10^{-3}$) and sensitivity (*FABP3, HIF1AN, PIK3R1, SCD*; *P*-value $=7.39\times10^{-3}$). In addition, interesting functions such as heart disease, blood pressure, glucose tolerance, synthesis of carbohydrate and glucose metabolism disorder, biogenesis of cholesterol, differentiation of muscle cells and adiposity were also identified (Table S7).

Finally, a total of nine direct and nine indirect networks were obtained with IPA (Table S8). The top direct network was associated with cell death and survival, cellular development, connective tissue development and function (Figure 3). It showed a score of 55 and contained 29 molecules (Table S8). The top indirect network was related to metabolic disease, lipid metabolism and molecular transport (Figure S1). A total of 21 molecules were associated to this network having a score of 36 (Table S8).

Discussion

To date, muscle transcriptome analyses concerning meat quality in swine have mainly been conducted using microarrays [3,27,30–32]. Compared with microarrays, RNA-Seq enables to determine the transcript abundance with a larger dynamic range of expression levels, it is not limited by the available genomic sequencing information during microarray production and can provide information about new isoforms. However, the main RNA-Seq drawback when compared with microarrays is that the analysis relies on the current pig genome assembly (in this study 10.2), in which interesting genes involved in lipid metabolism are still incorrectly annotated or not present. Therefore, the improvement of the annotation is transcendental for further RNA-Seq studies.

Muscle transcriptome description

In the present study, using RNA-Seq analysis we were able to map a high percentage of reads to the current pig genome assembly (Sscrofa10.2). Our percentage of mapped reads (85.1%) was similar to the described in the pig adipose tissue transcriptome (80%–87%) [5] performed with the Sscrofa10.2 annotation version, however it was higher than the percentage found in the pig muscle transcriptome (64,8%) [6] performed with the Sscrofa9.2 version or the pig liver transcriptome (71.4%–77.7%) [4] using the Sscrofa9.61 annotation version. The high amount of transcripts mapping to intergenic regions and the novel coding gene discovery, showing a higher percentage of computationally predicted proteins (60.2%) versus known proteins (39.8%) in pig in comparison to other species such as bovine and human, reinforces the need to improve the current pig annotation. Similar results were shown in the porcine liver [4] and adipose tissue [5] transcriptomes, in which the 86.0% and 62.5% of the novel proteins identified respectively were computationally-predicted. As expected, the major overlap of predicted novel proteins was between muscle and adipose tissue [5] because unlike the liver and gonads [2,4], both analyses were performed using the most recent annotation of the genome. Hence, a total of 40% novel predicted proteins in the muscle tissue transcriptome have also been found in adipose tissue, either realised with the Sscrofa10.2. Of the 2,372 predicted novel proteins, 972 were validated *in silico* being present in at least one of the three tissues compared [2,4,5]. Interestingly, 36 of the novel predicted proteins were also identified in four different tissues (liver [4], gonads [2], adipose [5], and muscle tissue) (Figure 4). When analysing the main metabolic pathways for the novel transcripts identified, the "phosphatidylinositol signalling system" and "inositol phosphate metabolism" were among the most represented categories. The phosphatidylinositol signalling system plays a critical role in the regulation of diverse processes such as muscle contraction, cell secretion, cell growth and differentiation [33]. Moreover, phosphatidylinositol is an essential component of the lipid membrane, where the total amount of phospholipids remains fairly constant, or increases little, as the animal increases in fatness [9,34]. Not surprisingly, these pathways were also identified when analyzing the adipose tissue novel transcripts [5]. Interestingly, the phosphatidylinositol signalling was also found within the most significantly overrepresented pathways in animals differing in FA composition in an study using microarrays [35]. Finally, we detected a high percentage (34.5%) of new repetitive elements present in the porcine genome. This result was similar to those obtained in adipose tissue (36%) using the Sscrofa10.2 genome annotation, but higher than those obtained in gonads (7.3%) and liver (approximately 5.8–7.3%)

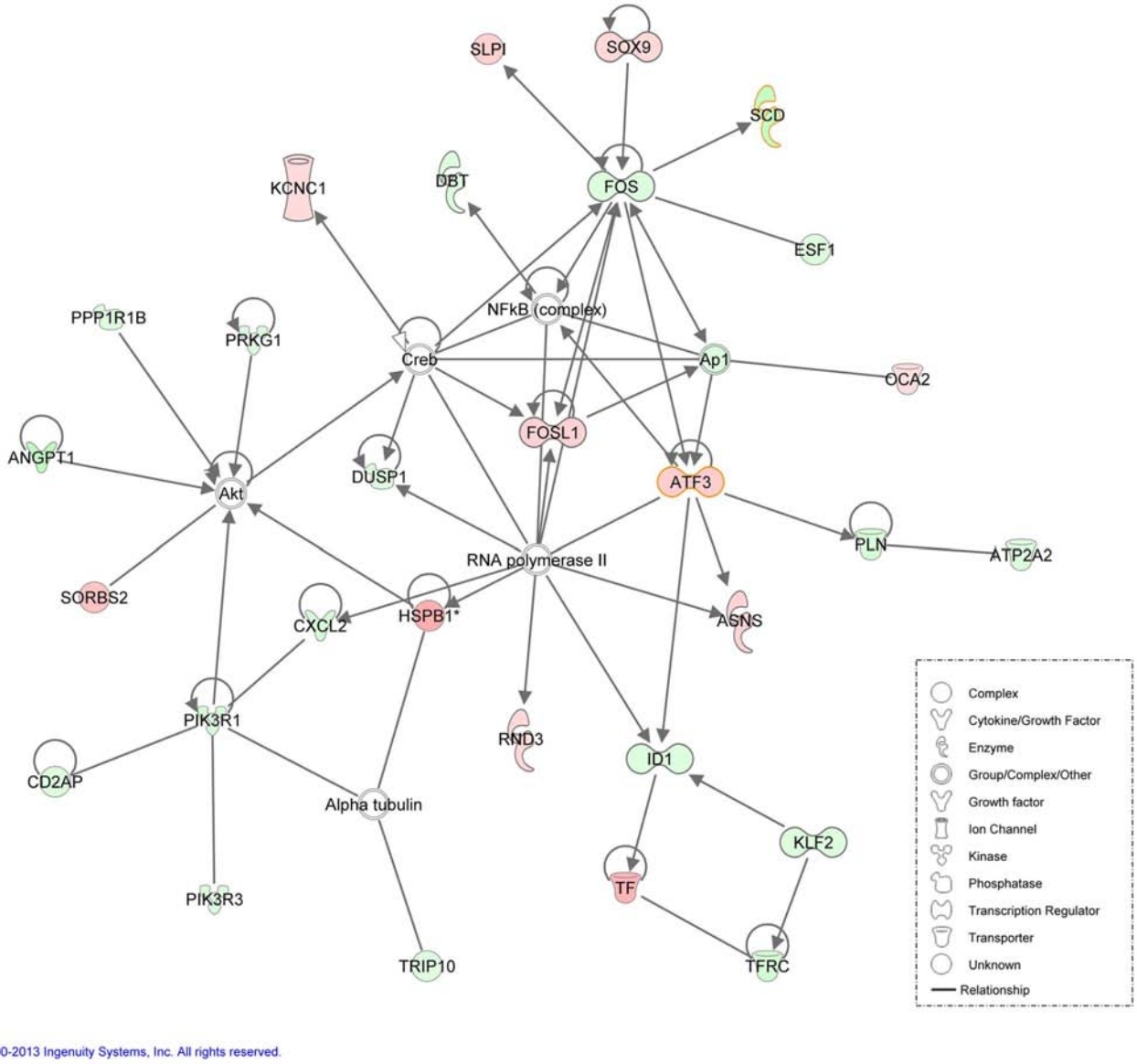


Figure 3. Network (direct, score 55) generated by IPA of 35 focus genes corresponding to the cell death and survival, cellular development, connective tissue and function pathways. Node colours indicate gene expression, being the red nodes higher-expressed genes and the green nodes lower-expressed genes in the H group relative to the L group. Colour intensity is related to the degree of expression. Node shapes indicate the biological function of the protein.
doi:10.1371/journal.pone.0099720.g003

that used an older version of the pig genome assembly [2,4,5]. This higher content of repetitive elements can be explained by the improvement of the current assembly (Sscrofa10.2) of the pig genome being the repetitive regions the most difficult to assemble [36].

Differential Expression analysis

Apart from describing the transcriptome of the *Longissimus dorsi* muscle, this study aimed to identify genes that can be implicated in determining the phenotypic differences of animals with extreme IMF FA composition belonging to an Iberian x Landrace backcross (BC1_LD). Iberian pigs are a local Mediterranean breed, and in comparison with Landrace, they have an extreme trend to obesity, with a higher IMF content and a strongest development of backfat tissue [8]. In contrast to the Iberian pigs, commercial breeds such as Landrace have suffered a strong selection towards a lean meat content, muscularity and enhanced reproduction [37]. Moreover, these two breeds are extreme for FA composition, showing the Iberian pigs a higher content of SFA and MUFA (specially C16 and C18:1) and the Landrace pigs a higher content of PUFA [8,38]. In our study, animals belonging to L group had a higher content of SFA and MUFA similarly to the Iberian pigs, whereas animals from H group had higher content of PUFA, as observed in the Landrace animals. Thus, this animal material suits very well to studies aiming at identifying the molecular factors influencing the FA metabolism in pigs.

For the differential-expression analysis we intersected the two lists of genes, obtained by DEseq [25] and EdgeR [24], to obtain a

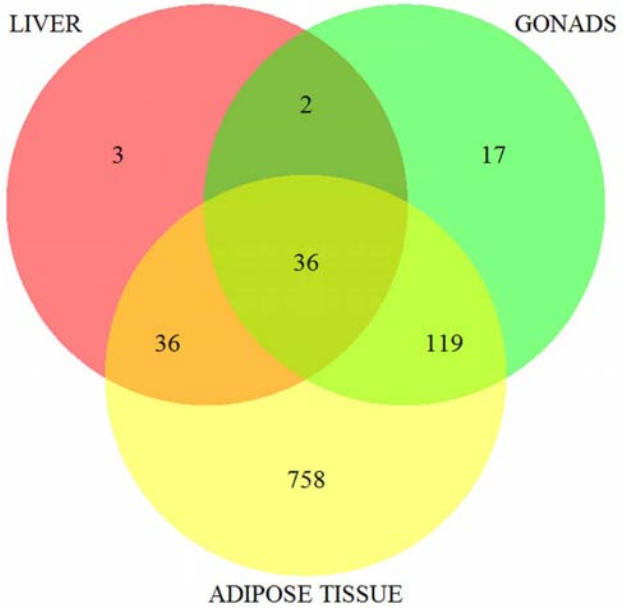


Figure 4. Venn diagram of the novel predicted proteins expressed in muscle, liver [4], gonads [2] and adipose tissue [5].
doi:10.1371/journal.pone.0099720.g004

single list in order to avoid false positives [39]. For the same reason we used a strict *P*-value ≤0.01, based on the Q-Q plot and equivalent to a FDR ≤0.12, and a fold change ≥1.2 as used in the adipose transcriptome analysis [5,40]. We identified a lower number of differentially-expressed genes with DESeq (208 genes) when compared to EdgeR (314 genes) what is in accordance with the observations reported by Soneson & Delorenzi [40]. DESeq uses conservative default settings and performs well when outliers are introduced, having a better false discovery rate (FDR) control for large sample sizes than EdgeR [40]. Notice that some relevant genes identified using microarrays in the muscle transcriptome of animals extreme for FA metabolism [27,35] such as *ACACA*, *FABP4* or *PPARGC1A* remained incorrectly or non-annotated in the Sscrofa10.2 annotation version. On the other hand, interesting genes detected in our RNA-Seq study that may determine differences in FA composition in muscle such as *ChREBP*, *GYS2*, *PLIN1*, *PLIN5* and *AQP7* could not be detected in microarray studies since probes for these genes were not included.

Differentially modulated metabolic pathways between groups

Among the top canonical pathways overrepresented between both groups of animals, we found hemostasis, nitric oxide (NO), metabolism of lipids and lipoproteins and PPAR and insulin signalling pathways (Table 2). Remarkably, most of the genes represented in these pathways were down-expressed in the H group. When compared with a previous study using microarrays of animals of BC1_LD population [27], the insulin and the calcium signalling, the regulation of the cytoskeleton, the focal adhesion dynamics, the leukocyte accumulation and cardiomyopathies-related pathways (Table S7) were found in common. Interestingly, our analysis identified other relevant pathways related to lipid metabolism, PPAR and NO signalling. On the other hand, most of the main overrepresented pathways identified in our study were also present in Duroc animals displaying divergent MUFA and PUFA fatty acids percentages (Table 2) [35], thus supporting a relevant role of these metabolic pathways in determining intramuscular FA composition. However, a feedback loop in which FA composition modifies these metabolic pathways that in turn cause a change in FA composition cannot be ruled out as we described below (i.e. differences in C16:1 n-7 FA or PUFA). Besides, in our analysis two other interesting pathways were found: (i) the clathrin-mediated endocytosis signalling, which is used for molecules such as low density lipoproteins, transferrins or growth factors and (ii) the C-X-C chemokine receptor type 4 (CXCR4) signalling, involved in the endocytosis of the glucose transporter protein 4 (GLUT4), specially in myocytes [41].

In the following sections a detailed explanation of differentially-expressed genes belonging to each of the main overrepresented pathways will be discussed.

- **NO and insulin signalling pathways.** The skeletal muscle is a target organ for the insulin-induced glucose uptake and for the maintenance of glucose homeostasis in blood [42]. Insulin acts in the carbohydrate metabolism facilitating the glucose diffusion into adipose and muscle cells via glucose transporter proteins (GLUT) and stimulates FA synthesis and the storage of triglycerides by the esterification of glycerol phosphate. Notably, the C16:1 n-7 FA, observed to be decreased in the H group (Table 1), can act as a lipokine that jointly with the expression of the *peroxisome proliferator-activated receptor gamma* (*PPAR-γ*) can strongly stimulate the muscle insulin action [43,44]. Interestingly, *PPAR-γ* and *solute carrier family 2, member 4* (*SLC2A4*; also called *GLUT4*) were identified as over-expressed in the L group animals when using EdgeR program. Supporting these results, Cánovas *et al.* [35] identified a higher expression of *myocyte enhancer factor 2A* (*MEF2A*) gene which upregulates *GLUT4* in Duroc animals having a higher MUFA and SFA content. Furthermore, insulin stimulates eNOS, the enzyme responsible for synthesizing NO by calcium-independent phosphorylation via phosphoinositide 3-kinases (PI3 kinases) and the downstream effector serine/threonine kinase (Akt) [45]. NO is a signalling molecule synthesized from L-arginine that plays an important role in regulating energy metabolism in mammals [46]. It has been reported that a chronic exposure of NO may decrease whole-body energy metabolism, increasing the adiposity and obesity [46]. For instance, *PI3 kinases*, down-expressed in the H group, have been reported to be necessary for the insulin-stimulated glucose uptake and glycogen synthesis, meanwhile, *Akt* regulates cell growth and metabolism and it is involved in glucose transport and lipogenesis (Figure 4) [16,47]. In the same direction, the *glycogen synthase* (*GYS*) gene was down-expressed in animals belonging to the H group (Figure S1), what might be a downstream effect of the Akt pathway [48]. Thus, the *GYS* inhibition may decrease the synthesis of glycogen necessary for glucose storage.
- **PPAR and metabolism of lipids and lipoproteins pathways.** In concordance with the low glucose-uptake that seems to occur in the H group of animals, we observed a down-expression of lipogenic genes most probably due to the lack of activation of carbohydrate responsive-element binding protein (*ChREBP*) [49]. The *stearoyl-CoA desaturase* (*SCD*) gene (Table S5) is responsible for the biosynthesis of MUFA from SFA, and its deficiency has been associated with lean mice [50]. Furthermore, polymorphisms in *SCD* gene have been strongly associated with FA composition in pigs and cows [12,51–53]. It has been suggested that an inhibition of this enzyme produces an increase in fatty acid oxidation through the inhibition of *acetyl-CoA carboxylase* (*ACACA*), regulated via

ChREBP, and *de novo* lipogenesis [50,54]. Interestingly, the *SCD* gene was identified as down-expressed in the adipose tissue of animals with higher content of PUFA in the Iberian x Landrace crossbred [5] and over-expressed in animals with higher IMF accumulation [31,35]. Our results support the hypothesis of Corominas *et al.* [5], that suggested that higher PUFA content in the H group suppresses the *ChREBP* gene function in a LXR-dependent manner inhibiting glycolytic and lipogenic genes. Although not present in the overlapping list (Table S5), the *ChREBP* gene was identified as down-expressed in the H group for EdgeR program.

Another gene whose disruption is associated with lean mice and was also down-expressed in H group is the *perilipin* [55]. Perilipins modulate the hydrolysis of triglycerides by *hormone-sensitive lipase* (*LIPE*) [56]. Specifically, *perilipin 5* (*PLIN5*) may play a role of "master lipolytic regulator" in muscle, and its over-expression can increase lipid droplet size and triacylglycerol storage [57]. We also identified the lipid transporter *Fatty acid binding protein 3, muscle and heart* (*FABP3*) and the *Aquoporins* (*AQP4* and *AQP7*) as down-expressed in the H group (Table S4). *FABP3* is more expressed in the skeletal muscle than in other tissues and participates in FA uptake and cytosolic transport, having a high binding affinity for palmitic, oleic and stearic acids. Furthermore, *FABP3* acts as a transcription factor in the nucleus for the control of lipid-mediated transcriptional programs via nuclear hormone receptors or other transcription factors that respond to lipids [58]. This gene has also been found in a genomic region significantly associated with FA composition in a GWAS performed in the Iberian x Landrace cross, being a clear candidate to explain the differences in FA composition observed between the two groups of animals [16]. Besides, it has been suggested as a candidate gene for the control of IMF deposition as it was identified as over-expressed in animals with higher IMF content [59]. The *Aquaporins* are modulated by the PI3K/Akt signalling and they are involved in glycerol uptake, particularly *AQP4* is localized in muscle fibers and it is important for energy supply in the skeletal muscle [46,60,61]. The *AQP7* which is higher expressed in fat tissue was also identified in the liver transcriptome study as being also down-expressed in animals with a higher content of PUFA [4,61]. Another differentially-expressed gene between the two groups of animals was the *very long-chain specific acyl-CoA dehydrogenase* gene (*ACADVL*), a *PPAR* target gene which was down-expressed in the H group. This gene catalyzes the first step of the mitochondrial FA β-oxidation pathway, mainly in muscle, having preference for C16:0, C16:1, C18:0 and C18:1 [62,63]. Moreover, *ACADVL* deficiency in humans produced a defective oxidation of oleic FA and knock-out mice for *ACADVL* fed in high-fat diet had a decrease in whole body fat content [64]. Overall, these results agree with a previous study in which the transcriptome of two groups of Duroc pigs with different IMF composition was analysed using microarrays and concluded that the IMF accumulation in animals having more IMF, MUFA and SFA may result from a balance between uptake, synthesis and degradation of triglycerides [35].

- **Hemostasis.** Alterations in fat metabolism play a role in the development of cardiovascular disease. Not surprisingly, our data set revealed several differentially-expressed transcripts which could be classified as potential regulators of hemostasis (Table S2). For instance, the *angiopoietin-1* (*ANGPT1*) gene which has been reported to increase vessel formation causing an enhanced glucose uptake and also the glycogen and lipid synthesis [65] was over-expressed in the L group and present in a QTL interval of the GWAS for IMF FA profile in the same population (Table S5) [16]. Furthermore, and consistent with our results, the angiogenesis promoted by ANGPT1 has been reported to increase NO production accompanied by an activation of the Akt and the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) signalling pathways [66,67] (Figure 4). Accordingly, the *hypoxia-inducible factor-1, alpha subunit inhibitor* (*HIF1AN*), which is regulated through the NF-κB inflammatory pathway and serves as an oxygen sensor regulating heart's oxygen supply, was down-expressed in the L group (Table S5). Thus, the L group animals having more SFA and MUFA content may have boosted the angiogenesis and improved the inflammatory response through the activation of the *ANGPT1* gene. A decreased essential PUFA content may lead to a proinflammatory eicosanoids synthesis and vasoconstrictors activation as has been reported in other studies [68]. In this direction, an over-expression of genes encoding for the inositol 1,4,5-triphosphate receptor 2 (ITPR2) protein which activates the release of Ca(2+) in the vessels acting as vasoconstrictor and aldosterone A (ALDOA) which increases blood pressure when activated by angiotensin [69,70] was observed in L group.

Pig lipid metabolism affected by intramuscular FA composition

In general, our results show that differences in FA composition may influence the lipid metabolism determining the phenotypic variation observed between the two groups of animals. In previous studies of our group, we observed that in liver [4], a high content of PUFA (H group phenotype) shifted the metabolism towards the FA oxidation; meanwhile, in adipose tissue [5] inhibited lipogenesis. Accordingly, in other studies analyzing the muscle transcriptome using microarrays a favored FA oxidation and a reduced fatty acid uptake, lipogenesis and triacylglycerol synthesis was generally observed in the group with higher intramuscular PUFA content [27,35]. In our RNA-Seq study in muscle we observed an inhibition of glucose uptake and lipogenesis in the H group, which would produce a decrease in the triglyceride storage. Noteworthy, in adipose and muscle transcriptome analysis, the *albumin* (*ALB*) gene was identified as over-expressed in animals having a higher content of PUFA (H group) [5]. The ALB is a long chain FA transporter that enhances FA mobilization affecting cellular uptake and also plays an antioxidant function in plasma. In adipose tissue, we hypothesized that ALB was supplying the FFAs used for the oxidation in liver in pre-slaughtering fasting conditions [5]. In the same direction, our results suggested that in muscle there is also an increased input of FFAs from blood and adipose tissue in order to fulfil the high-energy requirements in the H group. We hypothesize that animals having a high content of SFA and MUFA such as the Iberian pig, which is a rustic and slow-growing breed, may have an enhanced adaptation to fasting thanks to their high availability of muscle energy stores. Thus, selection towards a fast growth in commercial pigs such as Landrace, may have affected the ability to adapt to food disposal fluctuations [71].

Implications

In conclusion, the genes identified here as differentially-expressed between extreme animals, the pathways and the gene networks, contribute to understand the differences in gene regulation between the two groups differing in the muscle FA

composition. The functional analysis showed a different regulation of the lipid metabolism between groups, being more prone either to lipolisis or to lipogenesis depending on their FA composition. Moreover, the enrichment analysis showed that muscle plays a key role in energy metabolism, mainly in glucose and lipid metabolism, observing that animals having more PUFA, had a shift of the metabolism towards the lipolisis and also a lower glucose uptake. There are also evidences that the joint metabolism of the liver, adipose tissue and muscle may have an integrated role in determining the final FA composition of muscle. Therefore, the study of the muscle transcriptome may provide clues to decipher the genetic basis of meat quality traits. Moreover, this study may be of high relevance since FA composition in meat has important consequences in human health [72]. Besides, we observed that among the DE genes there was an overrepresentation of the obesity and the insulin resistance pathways. The results here exposed are particularly interesting because these diseases have a high prevalence and the pig has been described as a suitable biomodel for human lipid-related metabolic diseases [73].

Methods

Animal samples and phenotypes

The IBMAP population was originated by the cross of 3 Iberian boars (Guadyerbas) with 31 Landrace sows [74]. In this study we used 144 animals from the BC1_LD generation obtained by crossing five F1 boars with 26 Landrace sows. Animals were fed *ad libitum* with a cereal-based commercial diet (see [8] for diet details) and slaughtered at 179.8±2.6 days. Animal care and procedures were performed following national and institutional guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries). Samples of the *Longissimus dorsi* muscle were collected, snap frozen in liquid nitrogen and stored at −80°C until RNA extraction.

A Principal Component Analysis (PCA) was performed for characters related with the FA profile in muscle (see [4] for detailed description of this analysis). Twelve extreme animals were selected according to the first principal component (6 H and 6 L) [4]. Only females were taken into account in order to remove the sex effect on FA composition. The R language [75] was used to perform the statistical analysis of phenotypic mean comparison using a linear model.

RNA isolation

Total RNA was isolated from the *Longissimus dorsi* muscle of 12 samples with RiboPure Isolation of High Quality Total RNA (Ambion, Austin, TX, USA). Total RNA was quantified in a NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE, USA) and Qubit (Invitrogen, Carlsbad, CA, USA). RNA purity and integrity was checked employing a Bioanalyzer-2100 (Agilent Technologies, Inc., Santa Clara, CA, USA). All samples had a RNA Integrity Number (RIN) above 8.5.

Paired-end raw sequences (75 bp) were generated using a Hi-Seq 2000 instrument (Illumina, Inc., San Diego, CA, USA) in CNAG institute (Centro Nacional de Análisis Genómico, Barcelona, España).

Mapping and annotation

We ran FastQC [76] for the quality control. Indexed reads were then mapped to the reference pig genome version 10.2 (Sscrofa10.2) and the annotation database Ensembl Genes 67 [77] using TopHat v2.0.1 [19] with an allowance of two mismatches for each read. The resulting bam files containing the aligned sequences, were subsequently merged with Samtools [78]. Reads were annotated using the intersectBed option of BEDtools [79]. Cufflinks v2.0.2 program [20] was used to assemble the transcripts with a minimum of 10 reads per alignment. Finally, Samtools [78] was employed to compute descriptive statistics.

Gene expression quantification and differential-expression analysis

The number of reads mapping to each gene was determined with the *comp-counts* option in Qualimap v5.0 program [80]. We discarded those genes with a group mean less than 20 counts. We calculated the Pearson correlation coefficient between the mean expression values of the H and L group using the *cor.test* function of R. For the differential expression analysis we used DESeq [25] and EdgeR [24] packages implemented in R. We considered as differentially expressed between H and L groups those genes identified by both programs (DESeq and EdgeR) with a fold change ≥ 1.2 and P-value ≤ 0.01, the same parameters used in Corominas *et al.* [5], these case for both programs. FDR was calculated using the R package qvalue [81].

Transposable elements and orthology analysis

We used RepeatMasker version open-3.3.0 [21] with the rm-20120418 database in order to identify repetitive and transposable elements in the pig muscle transcriptome. We used "quick search" and "pig" species options and the Search Engine NCBI/RMBLAST.

Intergenic expressed regions not annotated in the Sscrofa10.2 version assembly were identified using Cuffcompare [82] and extracted using our own Python and R scripts. Novel putative proteins were predicted with Augustus program [22]. Afterwards, using Blast2GO [23], we mapped and annotated the novel protein coding genes. Using BLASTP option (*E*-value hit filter 1.00E-6, annotation cutoff 55, gene ontology (GO) weight 5 and HSP-hit coverage cutoff 0) we checked their orthology with already annotated proteins in *Homo sapiens*, *Bos taurus* and *Sus scrofa* protein databases. The InterProScan specific tool implemented in Blast2GO was employed to refine the functional annotations. With the GO Slim options we selected the relevant GO terms belonging to the cellular component, biological process and molecular function categories. Parameters were set to 10 for the seq filter and 20 for node score filter. Finally the ontology level was set to 3.

Gene ontologies and pathways

The Ingenuity Pathways Analysis software [29] and FatiGO tools from Babelomics 4.3 [28] were used to identify the most relevant biological functions and pathways in which the differentially-expressed genes (between the H and L groups) were involved. IPA, which uses its own private databases, allowed us to identify biological relevant information, identifying overrepresented pathways using the BH multiple testing correction (FDR) at P-value <0.05, and generating biological networks. For FatiGO, we used KEGG [83] and Reactome [84] databases setting the cut-off FDR <0.1. The Mouse Genome Database (MGD) [26] was used in order to identify how mutant alleles driven in mice for the 18 identified genes common in GWAS and RNA-Seq analysis affected the phenotype.

Data Availability

The full data sets have been submitted to NCBI Sequence Read Archive (SRA) under Accession SRP039424, Bioproject: PRJNA240057.

Supporting Information

Figure S1 Network (indirect, score 36) generated by IPA of 35 focus genes corresponding to metabolic disease, lipid metabolism and molecular transport. Node colours indicate gene expression, being the red nodes higher-expressed genes and the green nodes lower-expressed genes in the H group relative to the L group. Colour intensity is related to the degree of expression. Node shapes indicate the biological function of the protein.
(TIF)

Table S1 Percentage of reads mapped for each sample and their localization (exonic, intronic or intergenic) regarding the pig reference genome sequence.
(DOCX)

Table S2 Total number of assembled transcripts with cufflinks.
(DOCX)

Table S3 Description of the repetitive elements identified in the intergenic transcripts of the swine muscle transcriptome.
(DOCX)

Table S4 New predicted novel proteins with Augustus which have orthologous known genes identified with BLASTP option of Blast2GO.
(XLSX)

Table S5 Differentially-expressed genes identified among extreme groups (High and Low) for fatty acid composition in muscle.
(DOCX)

Table S6 Overrepresented categories identified with Babelomics and IPA for the differentially-expressed genes.
(XLSX)

Table S7 Specific functions table identified with IPA for the differentially-expressed genes.
(XLSX)

Table S8 Top networks identified with IPA from the differential expressed genes between High and Low animals.
(XLS)

Acknowledgments

We wish to thank Anna Castelló for RNA isolation and quantification.

Author Contributions

Conceived and designed the experiments: JMF MB. Performed the experiments: APO JMF MB. Analyzed the data: APO YRC MB JE. Contributed reagents/materials/analysis tools: JMF J. Corominas J. Casellas DPM JE YRC. Wrote the paper: APO JMF MB NH.

References

1. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562–578. doi:10.1038/nprot.2012.016.
2. Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C, et al. (2011) Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. BMC Genomics 12: 552.
3. Chen C, Ai H, Ren J, Li W, Li P, et al. (2011) A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. BMC Genomics 12: 448.
4. Ramayo-Caldas Y, Mach N, Esteve-Codina A, Corominas J, Castelló A, et al. (2012) Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics 13: 547. doi:10.1186/1471-2164-13-547.
5. Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Estellé J, Castelló A, et al. (2013) Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. Press.
6. Jung WY, Kwon SG, Son M, Cho ES, Lee Y, et al. (2012) RNA-Seq Approach for Genetic Improvement of Meat Quality in Pig and Evolutionary Insight into the Substrate Specificity of Animal Carbonyl Reductases. PLoS ONE 7: e42198. doi:10.1371/journal.pone.0042198.
7. Food and Agriculture Organization (FAO). Available: http://www.fao.org. Accessed 2014 May 20.
8. Serra X, Gil F, Pérez-Enciso M, Oliver M, Vázquez J, et al. (1998) A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Livest Prod Sci 56: 215–223. doi:10.1016/S0301-6226(98)00151-1.
9. Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, et al. (2008) Fat deposition, fatty acid composition and meat quality: A review. Meat Sci 78: 343–358. doi:10.1016/j.meatsci.2007.07.019.
10. Reiter SS, Halsey CHC, Stronach BM, Bartosh JL, Owsley WF, et al. (2007) Lipid metabolism related gene-expression profiling in liver, skeletal muscle and adipose tissue in crossbred Duroc and Pietrain Pigs. Comp Biochem Physiol Part D Genomics Proteomics 2: 200–206. doi:10.1016/j.cbd.2007.04.008.
11. Zhang W, Xiao S, Samaraweera H, Lee EJ, Ahn DU (2010) Improving functional value of meat products. Spec Issue 56th Int Congr Meat Sci Technol 56th ICoMST 15–20 August 2010 Jeju Korea 86: 15–31. doi:10.1016/j.meatsci.2010.04.018.
12. Maharani D, Park H-B, Lee J-B, Yoo C-K, Lim H-T, et al. (2013) Association of the gene encoding stearoyl-CoA desaturase (SCD) with fatty acid composition in an intercross population between Landrace and Korean native pigs. Mol Biol Rep 40: 73–80. doi:10.1007/s11033-012-2014-0.
13. FAO (2008) Fats and fatty acids in human nutrition. Report of expert consultation.
14. Zhang S, Knight TJ, Stalder KJ, Goodwin RN, Lonergan SM, et al. (2009) Effects of breed, sex and halothane genotype on fatty acid composition of triacylglycerols and phospholipids in pork longissimus muscle. J Anim Breed Genet 126: 259–268. doi:10.1111/j.1439-0388.2008.00782.x.
15. Madeira MS, Pires VMR, Alfaia CM, Costa ASH, Luxton R, et al. (2013) Differential effects of reduced protein diets on fatty acid composition and gene expression in muscle and subcutaneous adipose tissue of Alentejana purebred and Large White × Landrace × Pietrain crossbred pigs. Br J Nutr FirstView: 1–14. doi:10.1017/S0007114512004916.
16. Ramayo-Caldas Y, Mercadé A, Castelló A, Yang B, Rodríguez C, et al. (2012) Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. J Anim Sci 90: 2883–2893. doi:10.2527/jas.2011-4900.
17. Muñoz M, Rodríguez MC, Alves E, Folch JM, Ibañez-Escriche N, et al. (2013) Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data.
18. Revilla M, Ramayo-Caldas Y, Castelló A, Corominas J, Puig-Oliveras A, et al. (2014) New insight into the SSC8 genetic determination of fatty acid composition in pigs. Genet Sel Evol In Press.
19. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-seq. Bioinformatics 25: 1105–1111.
20. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511–515.
21. RepeatMasker version open-3.3.0. Available: http://www.repeatmasker.org. Accessed 2014 May 20.
22. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637–644.
23. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676. doi:10.1093/bioinformatics/bti610.
24. Robinson M, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25.
25. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.
26. Mouse Genome Informatics. Available: http://www.informatics.jax.org. Accessed 2014 May 20.
27. Pena RN, Noguera JL, Casellas J, Díaz I, Fernández AI, et al. (2013) Transcriptional analysis of intramuscular fatty acid composition in the longissimus thoracis muscle of Iberian × Landrace back-crossed pigs. Anim Genet: n/a–n/a. doi:10.1111/age.12066.
28. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, et al. (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res 38: W210–W213. doi:10.1093/nar/gkq388.
29. Ingenuity Pathways Analysis software. Available: www.ingenuity.com. Accessed 2014 May 20.

30. Yu K, Shu G, Yuan F, Zhu X, Gao P, et al. (2013) Fatty Acid and Transcriptome Profiling of Longissimus Dorsi Muscles between Pig Breeds Differing in Meat Quality. 9 1: 108–118. doi:doi:10.7150/ijbs.5306.
31. Hamill R, Aslan O, Mullen A, O'Doherty J, McBryan J, et al. (2013) Transcriptome analysis of porcine M. semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction. BMC Genomics 14: 453.
32. Zhao X, Mo D, Li A, Gong W, Xiao S, et al. (2011) Comparative Analyses by Sequencing of Transcriptomes during Skeletal Muscle Development between Pig Breeds Differing in Muscle Growth Rate and Fatness. PLoS ONE 6: e19774. doi:10.1371/journal.pone.0019774.
33. Putney JW, Bird GSJ (1993) The Inositol Phosphate-Calcium Signaling System in Nonexcitable Cells. Endocr Rev 14: 610–631. doi:10.1210/edrv-14-5-610.
34. Lagace TA, Ridgway ND (2013) The role of phospholipids in the biological activity and structure of the endoplasmic reticulum. Funct Struct Divers Endoplasmic Reticulum 1833: 2499–2510. doi:10.1016/j.bbamcr.2013.05.018.
35. Canovas A, Quintanilla R, Amills M, Pena R (2010) Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics 11: 372.
36. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, et al. (2005) Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comput Biol 1: e22. doi:10.1371/journal.pcbi.0010022.
37. Wilkinson S, Lu ZH, Megens H-J, Archibald AL, Haley C, et al. (2013) Signatures of Diversifying Selection in European Pig Breeds. PLoS Genet 9: e1003453. doi:10.1371/journal.pgen.1003453.
38. Estévez M, Morcuende D, Cava López R (2003) Physico-chemical characteristics of M. Longissimus dorsi from three lines of free-range reared Iberian pigs slaughtered at 90 kg live-weight and commercial pigs: a comparative study. Meat Sci 64: 499–506. doi:10.1016/S0309-1740(02)00228-0.
39. Yendrek C, Ainsworth E, Thimmapuram J (2012) The bench scientist's guide to statistical analysis of RNA-Seq data. BMC Res Notes 5: 506.
40. Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14: 91.
41. Antonescu CN, Díaz M, Femia G, Planas JV, Klip A (2008) Clathrin-Dependent and Independent Endocytosis of Glucose Transporter 4 (GLUT4) in Myoblasts: Regulation by Mitochondrial Uncoupling. Traffic 9: 1173–1190. doi:10.1111/j.1600-0854.2008.00755.x.
42. Kubota T, Kubota N, Kadowaki T (2013) The role of endothelial insulin signaling in the regulation of glucose metabolism. Rev Endocr Metab Disord 14: 207–216. doi:10.1007/s11154-013-9242-z.
43. Cao H, Gerhold K, Mayers JR, Wiest MM, Watkins SM, et al. (2008) Identification of a Lipokine, a Lipid Hormone Linking Adipose Tissue to Systemic Metabolism. Cell 134: 933–944. doi:10.1016/j.cell.2008.07.048.
44. Leonardini A, Laviola L, Perrini S, Natalicchio A, Giorgino F (2009) Cross-Talk between PPAR and Insulin Signaling and Modulation of Insulin Sensitivity. PPAR Res 2009. Available: http://dx.doi.org/10.1155/2009/818945.
45. Wilcox G (2005) Insulin and Insulin Resistance. Clin Biochem Rev 2: 19–39.
46. Dai Z, Wu Z, Yang Y, Wang J, Satterfield MC, et al. (2013) Nitric oxide and energy metabolism in mammals. BioFactors 39: 383–391. doi:10.1002/biof.1099.
47. Summers SA, Garza LA, Zhou H, Birnbaum MJ (1998) Regulation of Insulin-Stimulated Glucose Transporter GLUT4 Translocation and Akt Kinase Activity by Ceramide. Mol Cell Biol 18: 5457–5464.
48. Luo J, Sobkiw CL, Hirshman MF, Logsdon MN, Li TQ, et al. (2006) Loss of class IA PI3K signaling in muscle leads to impaired muscle growth, insulin response, and hyperlipidemia. Cell Metab 3: 355–366. doi:10.1016/j.cmet.2006.04.003.
49. Herman MA, Peroni OD, Villoria J, Schon MR, Abumrad NA, et al. (2012) A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. Nature 484: 333–338. doi:10.1038/nature10986.
50. Cohen P, Miyazaki M, Socci ND, Hagge-Greenberg A, Liedtke W, et al. (2002) Role for Stearoyl-CoA Desaturase-1 in Leptin-Mediated Weight Loss. Science 297: 240–243. doi:10.1126/science.1071527.
51. Jiang Z, Michal J, Tobey D, Daniels T, Rule D, et al. (2008) Significant associations of stearoyl-CoA desaturase (SCD1) gene with fat deposition and composition in skeletal muscle. Int J Biol Sci 4: 345–351.
52. Renaville B, Prandi A, Fan B, Sepulcri A, Rothschild MF, et al. (2013) Candidate gene marker associations with fatty acid profiles in heavy pigs. Meat Sci 93: 495–500. doi:10.1016/j.meatsci.2012.11.019.
53. Estany J, Ros-Freixedes R, Tor M, Pena RN (2014) A Functional Variant in the Stearoyl-CoA Desaturase Gene Promoter Enhances Fatty Acid Desaturation in Pork. PLoS ONE 9: e86177. doi:10.1371/journal.pone.0086177.
54. Benhamed F, Denechaud P-D, Lemoine M, Robichon C, Moldes M, et al. (2012) The lipogenic transcription factor ChREBP dissociates hepatic steatosis from insulin resistance in mice and humans. J Clin Invest 122: 2176–2194. doi:10.1172/JCI41636.
55. Martinez-Botas J, Anderson JB, Tessier D, Lapillonne A, Chang BH-J, et al. (2000) Absence of perilipin results in leanness and reverses obesity in Leprdb/db mice. Nat Genet 26: 474–479. doi:10.1038/82630.
56. Dalen KT, Dahl T, Holter E, Arntsen B, Londos C, et al. (2007) LSDP5 is a PAT protein specifically expressed in fatty acid oxidizing tissues. Biochim Biophys Acta BBA - Mol Cell Biol Lipids 1771: 210–227. doi:10.1016/j.bbalip.2006.11.011.
57. Bosma M, Sparks LM, Hooiveld GJ, Jorgensen JA, Houten SM, et al. (2013) Overexpression of PLIN5 in skeletal muscle promotes oxidative gene expression and intramyocellular lipid content without compromising insulin sensitivity. Biochim Biophys Acta BBA - Mol Cell Biol Lipids 1831: 844–852. doi:10.1016/j.bbalip.2013.01.007.
58. Furuhashi M, Hotamisligil GS (2008) Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. Nat Rev Drug Discov 7: 489–503. doi:10.1038/nrd2589.
59. Damon M, Wyszynska-Koko J, Vincent A, Hérault F, Lebret B (2012) Comparison of Muscle Transcriptome between Pigs with Divergent Meat Quality Phenotypes Identifies Genes Related to Muscle Metabolism and Structure. PLoS ONE 7: e33763. doi:10.1371/journal.pone.0033763.
60. Basco D, Blaauw B, Pisani F, Sparaneo A, Nicchia GP, et al. (2013) AQP4-Dependent Water Transport Plays a Functional Role in Exercise-Induced Skeletal Muscle Adaptations. PLoS ONE 8: e58712. doi:10.1371/journal.pone.0058712.
61. Rodríguez A, Catalán V, Gómez-Ambrosi J, García-Navarro S, Rotellar F, et al. (2011) Insulin- and Leptin-Mediated Control of Aquaglyceroporins in Human Adipocytes and Hepatocytes Is Mediated via the PI3K/Akt/mTOR Signaling Cascade. J Clin Endocrinol Metab 96: E586–E597. doi:10.1210/jc.2010-1408.
62. He M, Rutledge S, Kelly D, Palmer C, Murdoch G, et al. (2007) A new genetic disorder in mitochondrial fatty acid beta-oxidation: ACAD9 deficiency. Am J Hum Genet 81: 87–103.
63. Rakhshandehroo M, Knoch B, Müller M, Kersten S (2010) Peroxisome Proliferator-Activated Receptor Alpha Target Genes. PPAR Res 2010. Available: http://dx.doi.org/10.1155/2010/612089.
64. Chegary M, Brinke H te, Ruiter JPN, Wijburg FA, Stoll MSK, et al. (2009) Mitochondrial long chain fatty acid β-oxidation in man and mouse. Biochim Biophys Acta BBA - Mol Cell Biol Lipids 1791: 806–815. doi:10.1016/j.bbalip.2009.05.006.
65. Sung HK, Kim Y-W, Choi SJ, Kim J-Y, Jeune KH, et al. (2009) COMP-angiopoietin-1 enhances skeletal muscle blood flow and insulin sensitivity in mice. Am J Physiol - Endocrinol Metab 297: E402–E409. doi:10.1152/ajpendo.00122.2009.
66. Chen J-X, Stinnett A (2008) Ang-1 Gene Therapy Inhibits Hypoxia-Inducible Factor-1α (HIF-1α)-Prolyl-4-Hydroxylase-2, Stabilizes HIF-1α Expression, and Normalizes Immature Vasculature in db/db Mice. Diabetes 57: 3335–3343. doi:10.2337/db08-0503.
67. Oubaha M, Gratton J-P (2009) Phosphorylation of endothelial nitric oxide synthase by atypical PKCζ contributes to angiopoietin-1–dependent inhibition of VEGF-induced endothelial permeability in vitro. Blood 114: 3343–3351. doi:10.1182/blood-2008-12-196584.
68. Novgorodtseva T, Kantur T, Karaman Y, Antonyuk M, Zhukova N (2011) Modification of fatty acids composition in erythrocytes lipids in arterial hypertension associated with dyslipidemia. Lipids Health Dis 10: 18.
69. Xi Q, Adebiyi A, Zhao G, Chapman KE, Waters CM, et al. (2008) IP3 Constricts Cerebral Arteries via IP3 Receptor–Mediated TRPC3 Channel Activation and Independently of Sarcoplasmic Reticulum Ca2+ Release. Circ Res 102: 1118–1126. doi:10.1161/CIRCRESAHA.108.173948.
70. Bayorh MA, Ganafa AA, Emmett N, Socci RR, Eatman D, et al. (2005) Alterations in Aldosterone and Angiotensin II Levels in Salt-Induced Hypertension. Clin Exp Hypertens 27: 355–367. doi:10.1081/CEH-57423.
71. Merlot E, Vincent A, Thomas F, Meunier-Salaün M-C, Damon M, et al. (2012) Health and immune traits of Basque and Large White pigs housed in a conventional or enriched environment. animal 6: 1290–1299.
72. Webb EC, O'Neill HA (2008) The animal fat paradox and meat quality. 54th Int Congr Meat Sci Technol 54th ICoMST 10-15August 2008 Cape Town South Afr 80: 28–36. doi:10.1016/j.meatsci.2008.05.029.
73. Lunney JK (2007) Advances in Swine Biomedical Model Genomics. Int J Biol Sci 3: 179–184. doi:doi:10.7150/ijbs.3.179.
74. Pérez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, et al. (2000) A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross. J Anim Sci 78: 2525–2531.
75. Ihaka R, Gentleman R (1996) R: A Language for Data Analysis and Graphics. J Comput Graph Stat 5: 299–314. doi:10.2307/1390807.
76. FastQC. Available: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/.Accessed 2014 May 20.
77. Ensembl Genes 67. Available: http://www.ensembl.org/info/data/ftp/index.html.Accessed 2014 May 20.
78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079. doi:10.1093/bioinformatics/btp352.
79. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. doi:10.1093/bioinformatics/btq033.
80. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, et al. (2012) Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28: 2678–2679. doi:10.1093/bioinformatics/bts503.
81. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci 100: 9440–9445. doi:10.1073/pnas.1530509100.
82. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nat Methods 5: 621–628.
83. Kyoto Encyclopedia of Genes and Genomes. Available: http://www.genome.jp/kegg/pathway.html. Accessed 2014 May 20.
84. Reactome. Available: http://www.reactome.org. Accessed 2014 May 20.

A co-association network analysis of the genetic determination of pig conformation, growth and fatness

Puig-Oliveras A, Ballester M, Corominas J, Revilla M, Estellé J, Fernández AI, Ramayo-Caldas Y, Folch JM

PLoS One. 2014; 9(12): e114862. doi: 10.1371/journal.pone.0114862.

PLOS ONE

RESEARCH ARTICLE

A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness

Anna Puig-Oliveras[1,2]*, Maria Ballester[1,2], Jordi Corominas[1,2], Manuel Revilla[1,2], Jordi Estellé[3,4,5], Ana I. Fernández[6], Yuliaxis Ramayo-Caldas[1,2,3,4,5], Josep M. Folch[1,2]

1. Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra, Spain, **2.** Plant and Animal Genomics, Centre de Recerca en Agrigenòmica (CRAG), 08193, Bellaterra, Spain, **3.** Génétique Animale et Biologie Intégrative UMR1313 (GABI), Institut National de la Recherche Agronomique (INRA), 78350, Jouy-en-Josas, France, **4.** Génétique Animale et Biologie Intégrative UMR1313 (GABI), AgroParisTech, 78350, Jouy-en-Josas, France, **5.** Laboratoire de Radiobiologie et Etude du Génome (LREG), Commissariat à l'énergie atomique et aux énergies alternatives (CEA), 78350, Jouy-en-Josas, France, **6.** Departamento de Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040, Madrid, Spain

*anna.puig@cragenomica.es

OPEN ACCESS

Citation: Puig-Oliveras A, Ballester M, Corominas J, Revilla M, Estellé J, et al. (2014) A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. PLoS ONE 9(12): e114862. doi:10.1371/journal.pone.0114862

Editor: Marinus F.W. te Pas, Wageningen UR Livestock Research, Netherlands

Received: July 29, 2014

Accepted: November 14, 2014

Published: December 11, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work has been funded by MICINN project AGL2011-29821-C02 (Ministerio de Economía y Competitividad), and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036, Centre de Recerca en Agrigenòmica). APO was funded by a Personal Investigador en Formación (PIF) PhD grant from the Universitat Autònoma de Barcelona (458-01-1/2011). JC was funded by a Formación de Personal Investigador (FPI) PhD grant from Spanish Ministerio de Educación (BES-2009-018223), YRC by a Formación del Profesorado Universitario (FPU) PhD grant from the Spanish Ministerio (AP2008-01450), and MR by a Formació i Contractació de Personal Investigador Novell (FI-DGR) PhD grant from Generalitat de Catalunya (ECO/1639/2013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Background: Several QTLs have been identified for major economically relevant traits in livestock, such as growth and meat quality, revealing the complex genetic architecture of these traits. The use of network approaches considering the interactions of multiple molecules and traits provides useful insights into the molecular underpinnings of complex traits. Here, a network based methodology, named Association Weight Matrix, was applied to study gene interactions and pathways affecting pig conformation, growth and fatness traits.

Results: The co-association network analysis underpinned three transcription factors, *PPARγ*, *ELF1,* and *PRDM16* involved in mesoderm tissue differentiation. Fifty-four genes in the network belonged to growth-related ontologies and 46 of them were common with a similar study for growth in cattle supporting our results. The functional analysis uncovered the lipid metabolism and the corticotrophin and gonadotrophin release hormone pathways among the most important pathways influencing these traits. Our results suggest that the genes and pathways here identified are important determining either the total body weight of the animal and the fat content. For instance, a switch in the mesoderm tissue differentiation may determinate the age-related preferred pathways being in the puberty stage those related with the miogenic and osteogenic lineages; on the contrary, in the maturity stage cells may be more prone to the adipocyte fate. Hence, our results

demonstrate that an integrative genomic co-association analysis is a powerful approach for identifying new connections and interactions among genes.
Conclusions: This work provides insights about pathways and key regulators which may be important determining the animal growth, conformation and body proportions and fatness traits. Molecular information concerning genes and pathways here described may be crucial for the improvement of genetic breeding programs applied to pork meat production.

Introduction

About 43% of the meat consumed worldwide proceeds from pigs, thus representing the major source of meat for human food intake [1]. Moreover, pig serves as a model for metabolic diseases such as obesity in humans [2, 3]. For meat industry, carcass conformation and growth are economically important traits, determining the proportions of the different commercial cuts [4]. Understanding the interactions between genes defining body growth and conformation of pigs is therefore critical for an efficient pig production.

Over 553 quantitative trait loci (QTLs) for growth-related traits have been reported in pigs [http://www.animalgenome.org/cgi-bin/QTLdb/SS/index]. Moreover, a genome wide linkage analysis for growth and body composition carried out in an Iberian × Landrace cross (IBMAP) confirmed previous QTL regions and identified new ones in 10 of the 18 autosomes [5]. Despite the large number of QTLs identified by QTL scan and Genome-Wide Association Studies (GWAS) the genetic architecture of these complex traits is far from being understood [6]. The detection of SNPs having a clear effect on complex traits using GWAS is limiting, still being a challenging task. The main reason is because many genes have a little effect, moreover, the need for multiple tests correction methods may result in removing some interesting SNPs [7]. The power of single trait GWAS can be enhanced when considering simultaneously multiple phenotypes because complex traits generally have multiple correlated traits [7].

Hence, for complex traits, a systems biology approach that integrates the results into coherent network models offers many advantages over single trait approaches [8]. Recently, a framework for integrating the information of GWAS with network inference algorithms, named Association Weight Matrix (AWM), was developed to reveal and identify key regulatory elements, provide *in silico* information and generate gene networks with the aim to better understand the regulatory mechanisms of complex traits [9, 10]. However, few studies have been performed to date using system biology approaches and genotypic data in livestock species [9, 11–15].

Network biology approaches may substantially improve our knowledge about the diverse molecular pathways underlying complex traits. Using this methodology, the main objective of this work was to identify key regulators, gene

interactions and pathways determining pig growth and conformation traits in order to improve our knowledge about the architecture of these complex traits.

Results and Discussion

Global growth network description and trait cluster analysis

In the present study, we used a systems biology approach considering 12 growth-related phenotypes (Table 1). Given that primary cuts have economic impact in the Iberian pig production [16], ham weight was considered as the key trait for the AWM analysis. Among the genotypes of the 60K SNPs Porcine Beadchip, a total of 41,279 SNPs were retained for further analysis. Single-trait-single-SNP analysis by GWAS was performed for all traits (S1 Figure). The AWM approach captured a total of 1,747 annotated genes proximal to co-associated SNPs for conformation, growth and fatness traits. Therefore, an AWM with 1,747 nodes, representing genes, and a total of 316,166 edges, which account for the predicted interactions, was built (Fig. 1A). Interestingly, in the hierarchical cluster analysis two groups of phenotypic traits were formed showing a clear opposite directionality of the additive values. The first one containing the fatness traits (BFT155, BFT180, BFTS and IMF), whereas the second one encompassing the growth and conformation related traits (BW125, BW155, BW180, HW, SW, BLW, CW and CL) (S2 Figure). Additionally, a second cluster analysis considering only 54 genes (S1 Table) of the network which are known to be related with growth was performed, showing again, after clustering, two different groups for the additive values of growth and fatness traits (Fig. 2).

Next, in order to simplify and visualize the data with Cytoscape software, the number of interactions was reduced by selecting only the strongest co-associations, major than 0.86 ($\bar{X} \pm \sigma = 0.79 + 0.07$). The resulting network had 53,200 predicted interactions and 1,703 genes.

Key transcription factors regulating growth traits

Within the 1,703 associated-genes, a total of 142 putative regulators (S2 Table) were identified. After exploring all the possible interconnected trios among regulators, the top trio which spanned most of the network topology with highest connectivity (a total of 26,160 connections) and minimum redundancy was formed by the *Peroxisome Proliferator-Activated Receptor Gamma* (*PPARγ*; $PPAR\gamma_{Deg} = 147$), the *E74-Like Factor 1* (*Ets Domain Transcription Factor*) (*ELF1*; $ELF1_{Deg} = 237$), and the *PR Domain Containing 16* (*PRDM16*; $PRDM16_{Deg} = 256$) genes. In the resulting network, there were a total of 639 co-associations with the top trio of TF connecting 513 genes (Fig. 1B). Interestingly, *ELF1* localized in a QTL on SSC11 identified for growth and body composition traits in the IBMAP cross [5], whereas no QTL was identified on SSC6 and SSC13 regions where the two other TF, *PRDM16* and *PPARG,* were located. This result supports that the network methodology allowed the detection of potential variations affecting the

Table 1. Phenotypic traits registered in the BC1_LD (F1 × Landrace) and in the BC (F2 × Landrace) and F3 generations of the Iberian × Landrace cross.

Trait	Abbreviation	Statistics		
		N	Mean	SD
Body weight at 125 days (kg)	BW125	270	58.11	8.71
Body weight at 155 days (kg)	BW155	269	80.74	13.61
Body weight at 180 days (kg)	BW180	269	100.10	14.91
Carcass weight (kg)	CW	271	74.46	11.07
Carcass length (cm)	CL	261	81.86	6.23
Backfat thickness at 155 days (mm)	BFT155	269	13.34	3.08
Backfat thickness at 180 days (mm)	BFT180	220	15.60	3.21
Backfat thickness at slaughter (mm)	BFTS	237	23.26	6.13
Intramuscular fat percentage (%)	IMF	247	1.52	0.78
Weight of hams (kg)	HW	271	21.62	3.39
Weight of shoulders (kg)	SW	271	10.04	1.72
Weight of belly (kg)	BLW	276	7.33	1.14

doi:10.1371/journal.pone.0114862.t001

analyzed traits that would have not been detected by using single-trait based approaches (Table 2).

In the network, *PPARγ* gene, which is a key regulator of adipocyte differentiation, glucose homeostasis and fatty acid metabolism, was highly connected presenting 147 co-associations with other genes. This gene plays a role in determining the energy balance and the fat deposition influencing growth and body size [17, 18]. Furthermore, *PPARγ* has been associated with obesity, diabetes and atherosclerosis [19], and it has been identified, using the same methodology, as a key transcription factor regulating cattle puberty-related traits [9]. Interestingly, in another study of our group, *PPARγ* was identified as over-expressed in pigs having more MUFA and SFA versus pigs with high PUFA content [20]. Other studies in pigs suggested that *PPARγ* is an excellent target for determining growth and fat deposition traits at a certain age in pigs [21, 22].

On the other hand, *ELF1* gene is a major regulator of haematopoiesis and energy metabolism [23]. *ELF1* has also been described to trigger the *NF-κB* pathway activation involved in cell growth and differentiation and in lipid metabolism [24, 25]. Noteworthy, other members of the *ELF1* gene family (ETS transcription regulator family) are known to regulate adipocyte and osteoblast differentiation [26]. Remarkably, *FOXP3* which interacts with *ELF1* has been identified as a central transcription factor regulating IMF in cattle using the same methodology [13, 27].

Finally, *PRDM16* gene is involved in the differentiation of the brown adipose tissue, specifically in the switch between myogenic and adipogenic lineages [28]. *PRDM16* has been reported to control the myogenic cell fate into brown fat cells in mice, however, pigs lack in brown fat tissue [28, 29]. *PRDM16* can also function by triggering nervous and haematopoietic systems and participates in the regulation of the oxidative stress [30].

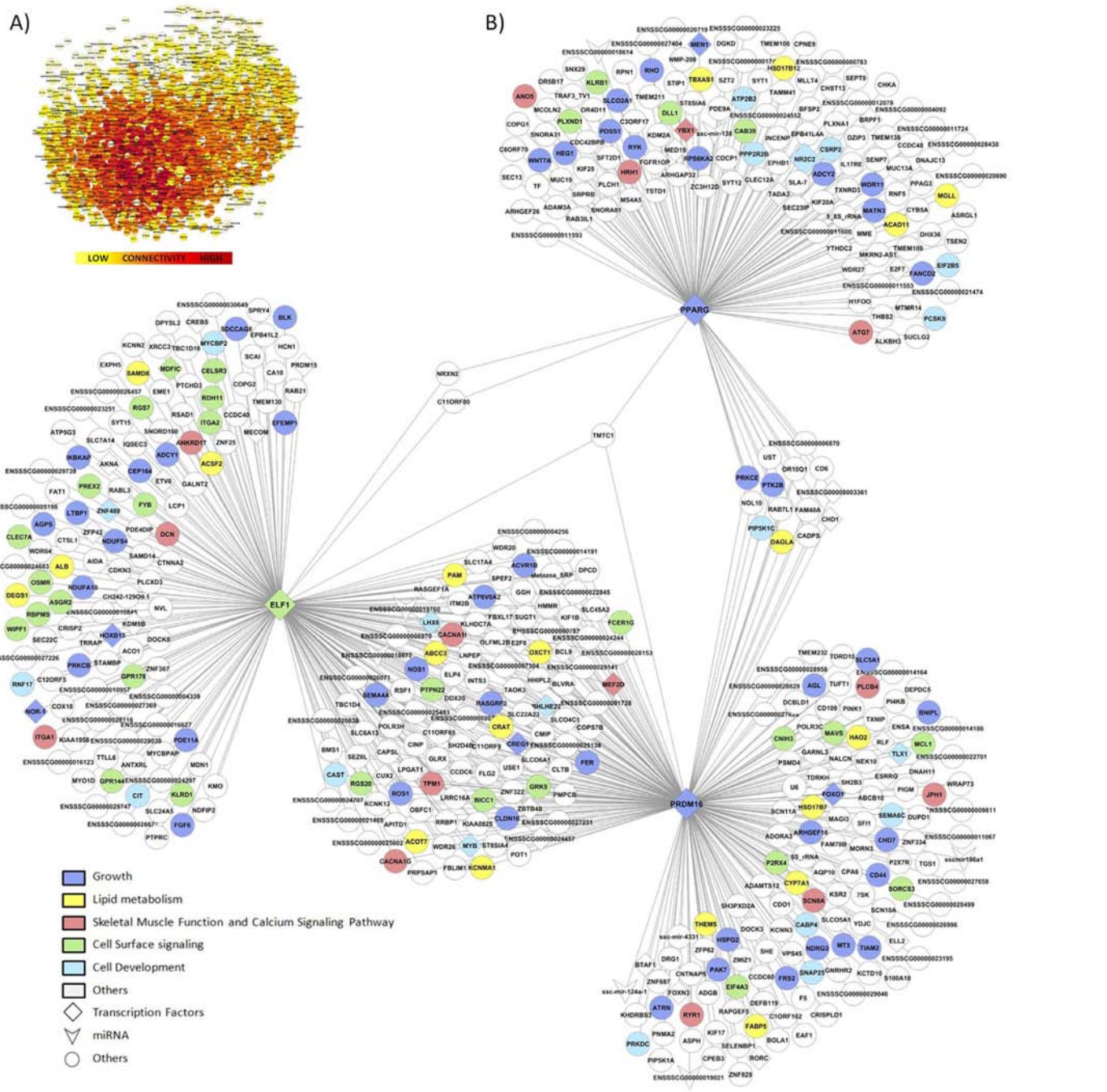


Fig. 1. Co-association networks based on the AWM approach. A) Full network formed by 1,747 nodes, representing genes and SNPs, and a total of 316,166 edges, accounting for the interactions among them. **B**) Network formed by 513 nodes and 639 edges representing genes and interactions among the top trio of transcription factors. Colours corresponded to different functions according to the legend.

doi:10.1371/journal.pone.0114862.g001

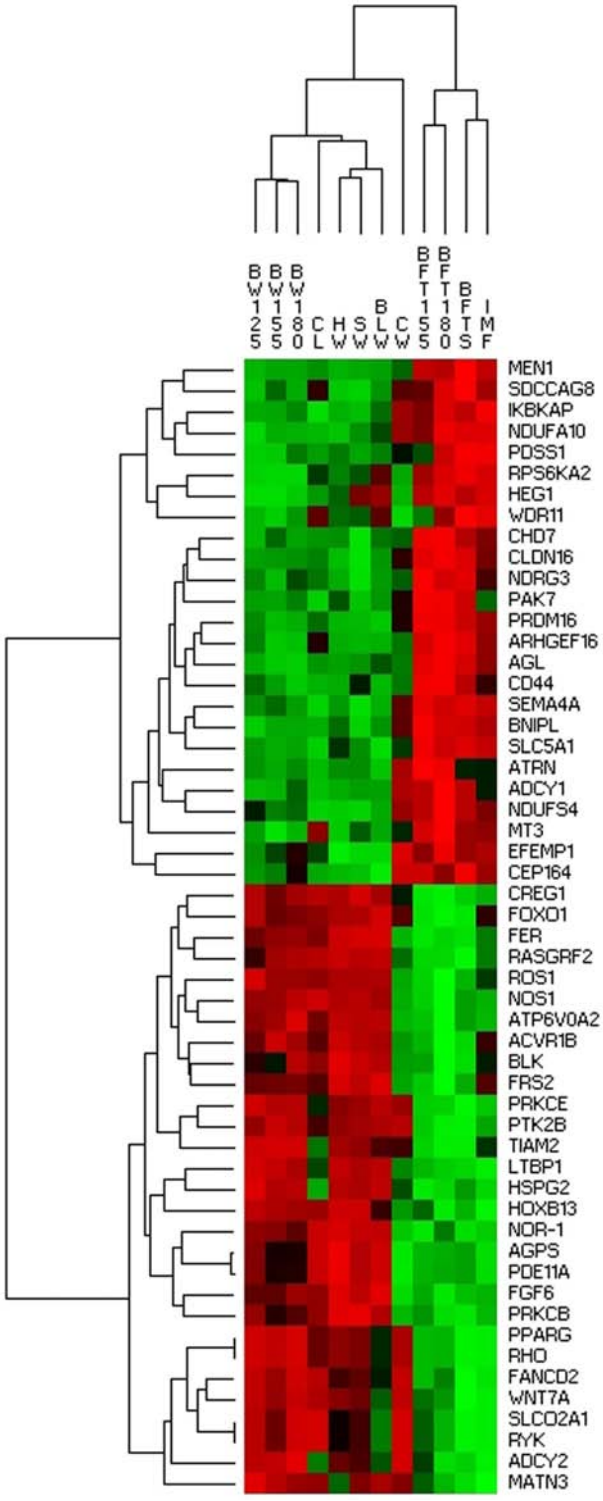


Fig. 2. Hierarchical cluster analysis considering only those genes in the network related with growth (S1 Table) among 12 phenotypic traits. The green colour in the figure corresponds to negative SNP additive effect values and red to positive SNP additive effect values.

doi:10.1371/journal.pone.0114862.g002

Table 2. Additive effect and p-value of the SNPs representing the top trio of transcription factors.

Gene	PRDM16		ELF1		PPARG	
Representative SNP	MARC0030882		MARC0000451		ISU10000701	
Trait	Additive Effect	P-value	Additive Effect	P-value	Additive Effect	P-value
BW125	−1.564	9.70E–02	1.767	2.89E–02	3.261	4.74E–05
BW155	−2.933	1.90E–02	2.217	4.09E–02	3.983	2.45E–04
BW180	−3.115	3.08E–02	1.762	1.59E–01	5.329	1.41E–05
CW	−0.734	5.23E–01	–0.008	9.95E–01	3.338	5.37E–04
CL	−0.534	1.28E–01	0.589	6.15E–02	0.758	1.26E–02
BFT155	0.630	4.78E–02	−0.011	1.00E+00	0.148	5.65E–01
BFT180	0.581	6.35E–02	−0.297	2.99E–01	0.161	5.54E–01
BFTS	0.666	3.38E–01	−0.526	3.46E–01	−0.537	3.73E–01
IMF	0.010	8.81E–01	−0.036	5.39E–01	−0.054	4.17E–01
HW	−0.580	1.16E–02	0.586	3.25E–03	0.546	6.01E–03
SW	−0.282	1.27E–02	0.229	2.19E–02	0.249	8.74E–03
BLW	−0.250	7.57E–03	0.174	3.24E–02	0.147	7.15E–02

doi:10.1371/journal.pone.0114862.t002

The embryonic mesoderm is a multipotent tissue that differentiates into myocytes, osteocytes and adipocytes [26]. The three top TF identified in the network have in common that are key regulators of the mesoderm cell fate. For instance, the over-expression of *PPARγ* may activate adipogenesis, *ELF1* may regulate adipocyte and osteoblast differentiation, meanwhile *PRDM16* may trigger the switch between adipose tissue and myocytes [28].

Selecting both miRNAs and TF as putative regulators did not affect the results, being the same top trio of genes identified as key regulators. In fact, inferring transcriptional and miRNA-mediated regulatory networks is still a challenge, particularly in non-model species such as the pig where the miRNA annotation is poor when compared to human or cow [31].

Additionally, a limitation of the AWM methodology is that only the nearest gene to the significant co-associated SNP is selected, discarding all other proximal genes. Linkage disequilibrium (LD) between molecular markers has to be taken into account for the AWM analysis. For instance, after exploring the network in more detail, a high co-association was observed between *Nuclear Receptor Subfamily 2, Group C, Member 2* (*NR2C2*) and *PPARγ* sharing the same co-associated nodes (Fig. 3). The strong relationship between *NR2C2* and *PPARγ* is supported by the literature, being *NR2C2* a repressor of *PPARγ* activity [32]. Interestingly, a deficiency of *NR2C2*, which has been suggested to play a critical role in the regulation of energy and lipid homeostasis, in mice causes growth retardation [32, 33]. Remarkably, *NR2C2* was also identified as co-associated in cattle growth network [12] and also in a network for fatness traits in cattle [13] and pig [11]. However, SNPs proximal to these genes (*PPARγ* and *NR2C2*) in our AWM network were separated by 1.27 Mb, being in complete LD ($D'=1$) (S3 Figure). Accordingly, LD can be a limitation to rule out which of these two genes play a key role regulating growth traits or if both genes are biologically relevant.

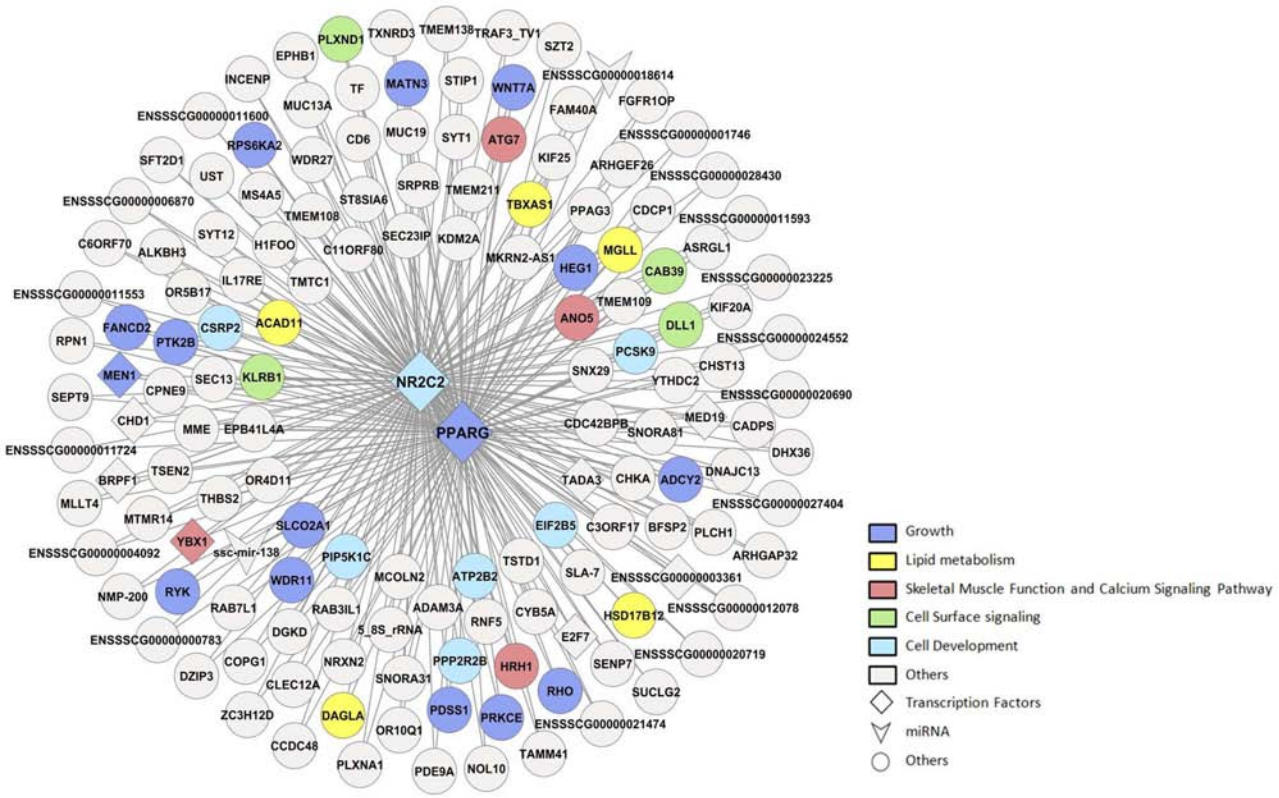


Fig. 3. Network showing the shared-genes of *NR2C2* and *PPARγ*.

doi:10.1371/journal.pone.0114862.g003

Co-association network among the top TF

The parameters describing the network topology were calculated with CentiScaPe software, obtaining an average degree (Deg) of 62.44 and an average distance (AvD_G) of 3.21; hence, showing a high degree of connection. A total of 54 genes out of the 513 nodes (Fig. 1B; blue colour in the network) belonged to growth-related gene ontologies (S1 Table). Noteworthy, a total of 20 genes were related to lipid metabolism (yellow colour in Fig. 1B). Among the 513 nodes setting up network connections, seven genes (*COPS7B, EFEMP1, ETV6, FRS2, HSPG2, SH3PXD2A* and *TGS1*) had been associated with human height, which is driven by growth and developmental processes [34, 35]. Interestingly, the *indian hedgehog* (*IHH*) gene product, identified in a human height GWAS study, binds to the patched domain containing 3 (PTCHD3) receptor, here identified as co-associated with *ELF1* [36]. In addition, 46 of the 513 genes were common with a study for cattle growth trait using the same methodology (*GRK5, NDRG3, RYK, FRS2, SCN8A, H1FOO, NALCN, EPB41L4A, LRRC16A, CNTNAP5, LTBP1, KHDRBS3, EPHB1, PRKCB, ATRN, TMEM108, PTK2B, RAPGEF5, RBPMS, SORCS3, SNX29, KCNN3, PLCH1, PLCB4, PDE11A, RGS7, NR2C2, WDR64, KCNMA1, DCN, SPEF2, CA10, MYB, RNF17, FYB, ETV6, CREB5, ZNF488, KSR2, SYT1, TBC1D16, SUCLG2, MLLT4, PLCXD3, VPS45, TUFT1*) [16]. Central to this

network, *transmembrane and tetratricopeptide repeat containing 1* (*TMTC1*) gene appeared to be a common interaction factor for the 3 principal TFs. Not *TMTC1* but *transmembrane and tetratricopeptide repeat containing 2* (*TMTC2*) was also identified in the cattle growth network by Widmann *et al.* [12].

Aside from known interactions reported by the literature, our growth network allowed the identification of new interactions between genes that have not been previously described and may help in the understanding of such complex trait. In these sense, one of the top TF identified in the growth network, *ELF1* gene, has not been reported to date to be involved in growth processes. This gene was identified to be co-associated with *B Lymphoid Tyrosine Kinase* (*BLK*) gene in the AWM analysis. It has been reported that *ELF1* is a transcriptional activator of *BLK* and *SRC* kinases such as *v-yes-1 Yamaguchi sarcoma viral related oncogene homolog* (*LYN*) [37]. *BLK* is involved in the stimulation of insulin secretion in response to glucose [38]. In addition, the *SRC* family of protein tyrosine kinases (*SFKs*) interacts with growth factors [39] and cytokine receptors [40] and they are key mediators of PI3K and AKT signalling important for cell proliferation [41]. The *LYN* gene belonging to *SFKs* is required for rapid phosphorylation of Fer (Fps/Fes Related) Tyrosine Kinase (FER) [42]. In the AWM analysis we found *FER* also co-associated with *ELF1*.

Another interesting interaction identified in the co-association growth network was the *ELF1* with *HOXB13*. Supporting this interaction, it has been described that Myeloid ecotropic viral integration site 1 (MEIS1) is a HOX cofactor which is regulated by *ELF1* [43]. *HOXB13* may play a role in growth repression and spinal cord formation [44, 45]. Furthermore, *HOXB13* increases the androgen and favors the lipid accumulation in cells [46].

The transcription factor *Forkhead box O1* (*FOXO1*) gene, also identified in the network co-associated with *PRDM16*, is activated in response to glucocorticoids and is blocked via the IGF/Akt pathway. *FOXO1* is a target of insulin signalling and glucose metabolism, as well as it plays a role in myogenic growth and differentiation. Moreover, it has been observed that mice overexpressing *FOXO1* in skeletal muscle had a reduced skeletal muscle mass when compared with wild-type mice [47]. When FOXO1 and PPARGC1A act together they promote gluconeogenesis. *FOXO1* is known to repress *PPARγ* [48]. Another interesting gene was the miRNA ssc-miR-196a1, which was identified as co-associated with *PRDM16*, and has been recently reported to be associated with growth and development of skeletal muscle [49]. Other identified co-associated genes with *PRDM16* were *SH3PXD2A, ADAMTS12* and *PTPN22*. Interestingly, SH3PXD2A (SH3 And PX Domains 2A) is reported to bind the matrix metalloproteinases (ADAMs) and phosphoinositides [50]. Moreover, in human is associated with *ADAM12* (the membrane-anchored protein corresponding to the secreted protein ADAMTS12) which is involved in skeletal muscle regeneration and mediates the neurotoxic effect of beta-amyloid peptide [51]. Finally, *PTPN22* has been associated with diabetes in humans [52].

Biological pathways and functional analysis

Functional analysis using IPA program allowed us to identify the biological functions overrepresented considering the 513 co-associated genes related to the top trio of TF. Among the networks identified with IPA there were: "cell signalling, nucleic acid metabolism, cell-to-cell signalling and interaction", "organism development, DNA replication, recombination and repair, and lipid metabolism" and "hereditary disorder, neurological disease and developmental disorder", all of them having a score =38, "cell-to-cell signalling and interaction, nervous system development and function, cellular assembly and organization" (score =33) and "cellular development, nervous system development and function, behaviour" (score =29) (S3 Table). Remarkably, the top molecular and cellular functions identified were: "post-translational modification" (p-value $=9.01 \times 10^{-5}$), "cell-to-cell signalling and interaction" (p-value $=1.26 \times 10^{-4}$), "molecular transport" (p-value $=1.26 \times 10^{-4}$), "cellular development" (p-value $=2.89 \times 10^{-4}$) and "cell morphology" (p-value $=3.10 \times 10^{-4}$). Among the top physiological system development functions we observed the "organismal survival" (p-value $=1.18 \times 10^{-4}$), "nervous system development and function" (p-value $=1.26 \times 10^{-4}$), "tissue development" (p-value $=2.34 \times 10^{-4}$) and "behaviour" (p-value $=2.88 \times 10^{-4}$). These results are in agreement with those obtained by Widmann *et al.* [12] using the same methodology to study growth traits in cattle, where they identified similar biological processes (cell communication, signal transduction, cellular process, cell surface receptor signalling pathway and cell adhesion) suggesting that different genetic variants may be affecting the same pathways even in different species.

Among the most overrepresented pathways identified we observed D-myo-inositol (1,4,5)-triphosphate ($Ins(1,4,5)P_3$) biosynthesis (p-value $=2.45 \times 10^{-4}$), G-protein coupled receptor (GPCR) signalling (p-value $=5.88 \times 10^{-3}$), corticotropin releasing (CRH) hormone signalling (p-value $=7.58 \times 10^{-3}$), gonadotropin-releasing hormone (GnRH) signalling (p-value $=1.55 \times 10^{-2}$), caveolar-mediated endocytosis signalling (p-value $=1.62 \times 10^{-2}$), nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation by viruses (p-value $=1.62 \times 10^{-2}$), phospholipase C (PLC) signalling (p-value $=1.94 \times 10^{-2}$) and neuronal nitric oxide synthase (nNOS) signalling in skeletal muscle cells (p-value $=3.46 \times 10^{-2}$) pathways (S4 Table). Noteworthy, in the cattle growth network study, the GnRH signalling and the nitric oxide (NO) pathway were also identified [16]. Some of these pathways (S4 Table) are discussed in more detailed below.

GPCR, PLC and $Ins(1,4,5)P_3$ signalling pathways

In mammals during growth and development there is a high requirement of lipids to increase in cell size and number. Lipids are the primary substrates which bind to certain GPCRs leading to an induced activity of PLC, which catalyse the hydrolysis of phosphatidylinositol 4,5-bisphosphate (PIP_2) to inositol 1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerol (DAG) both having important second messenger functions [53]. DAG can be also used as a component of biological

membranes or as a precursor to triacylglycerol (TAG) for energy storage [54]. The IP_3 molecule binds to the $Ins(1,4,5)P_3$ receptors (InsP3R) and trigger Ca^{2+} channel opening activating the ryanodine receptor-operated channel (RYR) [55, 56]. The IP_3 signalling mechanism is crucial for normal cell physiology [57]. Moreover, GPCRs jointly with phosphatidylinositol kinases (PIPK) may be involved in feed signal transduction pathways [58]. Two *PIPK* genes (*PIP5K1A* and *PIP5K1C*) jointly with *PLCB4* and *RYR1* were identified as co-associated to *PRDM16* gene. Thus, PRDM16 is a key transcriptional factor determining adiposity or miogenesis, and may also be necessary for the normal skeletal muscle development. Furthermore, the *GPCRs, GPR144* and *GPR176* were also identified in the growth network.

GnRH and CRH signalling pathway

Two hormone-related pathways, GnRH and CRH signalling pathways, were identified as overrepresented in our data. CRH is a peptide hormone secreted by the hypothalamus which controls adrenal secretion of cortisol and has been suggested to play a role in cell growth and survival. Interestingly, sheep with high cortisol response were prone to obesity [59]. Moreover, it has been observed that children treated with glucocorticoids showed growth retardation [60]. GnRH, which binds to GNRHR (GPCR member 7), is synthesized and released from neurons within the hypothalamus and regulates the production of gonadotropins, such as luteinizing hormone (LH) and follicle-stimulating hormone (FSH) in the pituitary gland, which, in turn triggers sexual maturation and promotes the secretion of endogenous sex hormones such as testosterone and estrogen from the gonads. Interestingly, growth-related traits in pig depend largely on gender [61]. In fact, GnRH agonists are used to treat central precocious puberty (CPP) characterized for developing an early puberty, larger growth of the skeleton and adult height [62]. Besides, the GnRF analog-diphtheria toxoid conjugate is used for castration and it is known to increase body weight at slaughter and improve average daily gain and feed conversion ratio [63]. The release of Ca^{2+} and DAG enhances the activation of protein kinases (PKC) to increase gonadotropin hormone secretion. Interestingly, in the network we can observe *PRKCB* co-associated to *ELF-1* and *PRKCE* co-associated with *PPARγ* and *PRDM16*. These results underline an important function of the three key TF in our network having an important role in the puberty and bone growth development.

NF-κB and NO signalling pathway

NF-κB is a pleiotropic transcription factor being involved in many biological processes such as inflammation, immunity, differentiation, cell growth and apoptosis. This signalling pathway has also been identified in a study of human growth using gene expression data [64]. It is reported that *ELF1* interacts with *NF-κB* via DNA binding domain [37]. Furthermore, the NF-κB activity may activate nNOS to generate NO [65]. NO may trigger GH secretion and affects other several pituitary peptides such as gonadotropins [66]. It is reported that a chronic exposure of NO may stimulate angiogenesis and adipocyte development [67].

Interestingly, we observed *NOS1* co-associated with *ELF1* in the growth network (Fig. 1B). Both, *ELF1* and *NOS1* genes may play a role in hematopoiesis and vascular development [68]. On the other hand, NO has the ability to enhance and regenerate diseased muscle, through determining the fibro-adipogenic progenitors fate and inhibiting adipogenesis [69]. In this direction, the NO production is thought to inhibit *PPARγ* expression [69]. Surprisingly, *NOS1* was also identified as co-associated with *PRDM16*, but not with *PPARγ*.

General discussion: from network to phenotype inference

Iberian pigs are known to have higher IMF than Landrace pigs at the same growth stage [70]. Furthermore, the skeletal muscle grows faster in Landrace than in Iberian pigs, being less prone to obesity [70]. In this study, we analyzed backcrossed animals from an Iberian × Landrace cross, which differed in fat and growth traits. Growth refers to an increase in tissue mass and it can be plotted as a sigmoid curve depending on age and cumulative weight [71]. At the pre-mature phase, muscle mass, organ and bone formation are increased, meanwhile in the mature phase the animal is more prone to fattening and intramuscular deposition [71]. The top trio of TF identified in the network are key regulators for mesoderm cell differentiation in osteocytes, miocytes or adipocytes. Our results showed among the top molecular and cellular functions the cell development and interaction pathways which may be important in order to trigger tissue formation. Noteworthy, the hierarchical cluster analysis evidenced a clear division of the additive effects of the SNPs for the 12 growth phenotypic traits, between animal weight-related and fat-related traits (S2 Figure and Fig. 2). We hypothesize that the hormone releasing pathways here identified (GnRH and CRH) may be key for the regulation of conformation, growth and fatness traits in our animal material. An increased carcass weight with a reduced backfat thickness at a fixed age have been selection targets in commercial pig breeds, resulting in less mature animals as fat deposition rate is expected to increase in the puberty phase [72]. Given that Iberian breed is more prone to IMF and backfat deposition at the same growth stage lead to the hypothesis that Landrace animals may arrive later to the mature growth phase when compared to Iberian animals. What remains unclear is whether the signals for maturity switching are related to the GnRH or CRH hormone release pathways. Finally, the genes and pathways here identified had a high concordance with those reported by other authors studying growth metabolism in animals or height related traits in human. Our hypothesis are supported by the high concordance between our genes and pathways identified in the network and those reported by Fortes *et al.* [9] for puberty traits in cattle.

Conclusions

The processes regulating conformation, growth and fatness traits in pigs are complex and most of the mechanisms remain unknown despite being of great interest for the pig industry. The power of single trait GWAS can be enhanced

when considering simultaneously multiple phenotypes taking advantage of system biology approaches. In the present study, the AWM gene co-association network analysis revealed key transcription factors, gene-gene interactions and pathways underpinning the regulation of pig conformation, growth and fatness. Network approaches represent a major step in understanding the genetics of complex diseases and traits. Further efforts should be made in order to study in more detail the new gene-gene interactions here identified, as well as, to study in more detail the key transcription factors and pathways involvement in the growth and conformation traits determination.

Material and Methods

Animal material and phenotypic classification

The animal material used belongs to several generations of the IBMAP population obtained from the cross of 3 Iberian boars (Guadyerbas) with 31 Landrace sows [73, 74]. For this study we used phenotypic records from 292 animals belonging to three different IBMAP generations: 159 BC1_LD animals (25% Iberian $\times$75% Landrace) from backcrossing five F1 males with 26 Landrace sows, 79 BC animals obtained by crossing 4 F2 boars and 22 Landrace sows and 54 F3 obtained by mating F2 animals. Animals were fed *ad libitum* and sacrificed at 180$\pm$2.8 days (average $\pm$ standard deviation) in a commercial slaughterhouse following national and institutional guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries).

Phenotypic records used in the analyses (Table 1) correspond to body weight (BW) measured at 125, 155 and 180 days (BW125, BW155, and BW180, respectively), backfat thickness (BFT) at the level of the fourth rib at 4 cm of the midline measured by ultrasounds at 155 and 180 days (BFT155 and BFT180) and measured with a rule at slaughter (BFTS), carcass length (CL) and carcass weight (CW), ham weight (HW), shoulder weight (SW), belly weight (BLW), and the intramuscular fat content (IMF) in the *longissimus dorsi* muscle.

Genetic markers and quality control

A total of 364 pigs, including their F0, F1 and F2 founder generations (72 animals), were genotyped with the Porcine SNP60K BeadChip [75] following the Infinium HD Assay Ultra protocol (Illumina Inc.; San Diego, CA, USA) and the genotypes were visualized with the GenomeStudio software (Illumina Inc.; San Diego, CA, USA). The quality control of the 62,163 SNPs was performed by using Plink [76] software removing markers with a minor allele frequency (MAF) $<$5% and animals with missing genotypes$>$5%. The SNP mapping and annotation was performed by using the pig assembly 10.2 [ftp://ftp.ncbi.hlm.nih.gov/genomes/Sus_scrofa/GEF/]. We also excluded markers which did not map in the

Sscrofa10.2 version assembly. Pedstats program [77] was used to check Mendelian inheritance errors.

Genome-wide association analysis

Genome-wide association analysis (GWAS) for the twelve phenotypic growth traits were performed using a mixed model accounting for additive effects with Qxpak 5.0 software [78]:

$$y_{ijlkm} = Sex_i + Batch_j + \beta c_1 + \lambda_1 a_k + u_1 + e_{ijlkm},$$

in which y_{ijlkm} was the i-th individual record, sex (two levels) and batch (nine levels) were fixed effects, ß was a covariate coefficient with *c* being the covariate used in each case (described below), λ_l was a −1, 0, +1 indicator variable depending on the l-th individual genotype for the k-th SNP, a_k represented the additive effect associated with the k-th SNP, u_l represented the infinitesimal genetic effect with random distribution $N(0, A\sigma_u)$ where **A** was a numerator of the the pedigree-based relationship matrix and e_{ijlkm} the residual.

Different covariates (*c*) were used for the analysis. Carcass weight was used as a covariate for CL, IMF, BFTS, HW, SW, and BLW. For BFT155 and BFT180 the covariates used were the body weight at their respective days. Meanwhile, for the body and the carcass weights the covariate used was the animal age.

Association weight matrix

The association weight matrix (AWM) was built from the GWAS results. First, the SNP additive effects were normalized with a z-score method using a R script and a matrix was constructed with these values, being SNPs in rows and traits in columns. Another matrix with the same format was generated for the p-values obtained in the GWAS. For the analysis, the ham weight was selected as the key phenotype. Subsequently, the AWM script [9] available from authors was used in R (http://www.r-project.org/). Those SNPs associated (nominal p-value <0.05) with the ham weight or with 3 or more traits were selected for further analysis. We included in the analysis the SNPs with a distance of minor than 2.5 kb (SNPs close) and major than 1,000 kb (SNPs far) from a gene. We also included SNPs located at less than 10 kb of miRNA. Finally, to facilitate the analysis, for SNPs clustering at less than 1 Mb of distance from each other, the SNP associated with the major number of characters was selected. The hierarchical clustering option of PermutMatrix software [79] was used to visualize the results of both traits and genes. The trio of putative regulators spanning most of the network topology with a minimum redundancy [10] was selected. In this study we took into account all the transcription factors (TF) from the list reported by Vaquerizas *et al.* [80]; additionally, those 22 genes belonging to the GO: 0050789 which accounts for the DNA binding TF activity were added. All miRNA annotated on Sscrofa10.2 assembly were also included in the analysis as potential regulators. PCIT

algorithm [81] was used to construct a file containing the reported gene-gene interactions among the 3 TFs. The CentiScaPe plug-in [82] of Cytoscape software [83] was used to visualize the PCIT results and either to calculate the node centrality values (Deg) and network parameters.

Gene ontologies, pathways and network analysis

The Ingenuity Pathways Analysis software (IPA; Ingenuity Systems, Redwood city, CA, USA; www.ingenuity.com) was used to identify the most relevant biological functions and pathways in which the genes associated with the phenotypic traits were involved. IPA, which uses its own databases, allowed the identification of overrepresented pathways using the BH multiple testing correction [84] of p-value (FDR <0.05) and generating biological networks. The Mouse Genome Database (MGD; http://www.informatics.jax.org) was used in order to identify how mutant alleles driven in mice for the identified growth-related genes present in the network affected the phenotype.

Supporting Information

S1 Figure. GWAS plot of the 12 traits: body weight measured at 125, 155 and 180 days (BW125, BW155, and BW180, respectively), backfat thickness measured at 155 and 180 days (BFT155 and BFT180) and measured at slaughter (BFTS), carcass length and weight (CL and CW), weight of the hams, shoulders and belly (HW, SW and BLW) and intramuscular fat (IMF) content. The horizontal green line represents the statistical significance (false discovery rate; set at q-value ≤ 0.05) calculated with the q-value library [85] implemented in R program (http://www.r-project.org/).
doi:10.1371/journal.pone.0114862.s001 (DOCX)

S2 Figure. Hierarchical cluster analysis among 12 phenotypic traits: body weight measured at 125, 155 and 180 days (BW125, BW155, and BW180, respectively), backfat thickness measured at 155 and 180 days (BFT155 and BFT180) and measured at slaughter (BFTS), carcass length and weight (CL and CW), weight of the hams, shoulders and belly (HW, SW and BLW) and the intramuscular fat (IMF) content.
doi:10.1371/journal.pone.0114862.s002 (TIF)

S3 Figure. Linkage disequilibrium among the *PPARG* and *NR2C2* SNPs. Pattern of linkage disequilibrium analysis around $\pm$2Mb of the SNPs in *PPARG* and *NR2C2*. Figure colored from blue to red according to LD strength between consecutive markers. The green diamond-shape corresponds to the SNP in *PPARG* gene and the blue diamond-shape the SNP in *NR2C2* gene.
doi:10.1371/journal.pone.0114862.s003 (DOCX)

S1 Table. List of 54 growth-related genes in the network.
doi:10.1371/journal.pone.0114862.s004 (XLSX)

S2 Table. List of 142 regulators (transcription factors and miRNAs) identified within the list of associated-genes.
doi:10.1371/journal.pone.0114862.s005 (XLSX)

S3 Table. Top networks of molecular functions identified with IPA for the 513 genes.
doi:10.1371/journal.pone.0114862.s006 (XLS)

S4 Table. Top pathways identified with IPA for the 513 genes.
doi:10.1371/journal.pone.0114862.s007 (XLS)

Acknowledgments

The authors gratefully acknowledge J.L. Noguera (Institut de Recerca i Tecnologia Agroalimentàries; IRTA) for the animal material contribution.

Author Contributions

Conceived and designed the experiments: APO JMF MB. Performed the experiments: APO. Analyzed the data: APO YRC. Contributed reagents/materials/analysis tools: JMF MB JC YRC AIF MR JE. Wrote the paper: APO JMF MB.

References

1. **FAO** (2008) Fats and fatty acids in human nutrition. Report of expert consultation.
2. **Dodson M, Hausman G, Guan L, Du M, Rasmussen T, et al.** (2010) Lipid metabolism, adipocyte depot physiology and utilization of meat animals as experimental models for metabolic research. Int J Biol Sci 6: 691–699.
3. **Walters E, Wolf E, Whyte J, Mao J, Renner S, et al.** (2012) Completion of the swine genome will simplify the production of swine as a large animal biomedical model. BMC Med Genomics 5: 55.
4. **Gispert M, Font i Furnols M, Gil M, Velarde A, Diestre A, et al.** (2007) Relationships between carcass quality parameters and genetic types. Meat Sci 77: 397–404. doi: 10.1016/j.meatsci.2007.04.006.
5. **Fernandez A, Perez-Montarelo D, Barragan C, Ramayo-Caldas Y, Ibanez-Escriche N, et al.** (2012) Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. BMC Genet 13: 41.
6. **McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al.** (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356–369. doi: 10.1038/nrg2344.
7. **Rao DC** (2008) An Overview of the Genetic Dissection of Complex Traits. In: D. C Rao and C. Charles Gu, editor. Advances in Genetics. Academic Press, Vol. Volume 60. pp. 3–34. Available: http://www.sciencedirect.com/science/article/pii/S0065266007004014.
8. **Wang K, Li M, Hakonarson H** (2010) Analysing biological pathways in genome-wide association studies. Nat Rev Genet 11: 843–854. doi: 10.1038/nrg2884.
9. **Fortes MRS, Reverter A, Zhang Y, Collis E, Nagaraj SH, et al.** (2010) Association weight matrix for the genetic dissection of puberty in beef cattle. Proc Natl Acad Sci 107: 13642–13647. doi: 10.1073/pnas.1002044107.
10. **Reverter A, Fortes MS** (2013) Association Weight Matrix: A Network-Based Approach Towards Functional Genome-Wide Association Studies. In: Gondro C, van der Werf J, Hayes B, editors. Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology. Humana Press, Vol. 1019. pp. 437–447. Available: http://dx.doi.org/10.1007/978-1-62703-447-0_20.

11. **Ramayo-Caldas Y, Ballester M, Fortes M, Esteve-Codina A, Castello A, et al.** (2014) From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. BMC Genomics 15: 232.

12. **Widmann P, Reverter A, Fortes MR, Weikard R, Suhre K, et al.** (2013) A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. BMC Genomics 14: 798.

13. **Ramayo-Caldas Y, Fortes MRS, Hudson NJ, Porto-Neto LR, Bolormaa S, et al.** (2014) A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G and FOXP3in intramuscular fat deposition of beef cattle. J Anim Sci. Available: http://www.journalofanimalscience.org/content/early/2014/04/28/jas.2013-7484.abstract.

14. **Fortes MRS, Reverter A, Nagaraj SH, Zhang Y, Jonsson NN, et al.** (2011) A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. J Anim Sci 89: 1669–1683. doi: 10.2527/jas.2010-3681.

15. **Fortes MRS, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, et al.** (2012) Gene network analyses of first service conception in Brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. J Anim Sci 90: 2894–2906. doi: 10.2527/jas.2011-4601.

16. **Fernández A, García-Casco J, De Pedro E, Silió L, Rodríguez MC** (2007) Genetic antagonism between intramuscular fat content and primal cuts in Iberian pigs? In: Casabianca F., Monin G., Audiot A., editors. 5. International Symposium on the Mediterranean Pig. Options Méditerranéennes: Série A. Séminaires Méditerranéens. Zaragoza: CIHEAM, Vol. 76. pp. 43–46. Available: http://om.ciheam.org/om/pdf/a76/00800557.pdf.

17. **Cecil JE, Fischer B, Doney ASF, Hetherington M, Watt P, et al.** (2005) The Pro12Ala and C–681G variants of the PPARG locus are associated with opposing growth phenotypes in young schoolchildren. Diabetologia 48: 1496–1502. doi: 10.1007/s00125-005-1817-0.

18. **Rieusset J, Seydoux J, Anghel SI, Escher P, Michalik L, et al.** (2004) Altered Growth in Male Peroxisome Proliferator-Activated Receptor γ (PPARγ) Heterozygous Mice: Involvement of PPARγ in a Negative Feedback Regulation of Growth Hormone Action. Mol Endocrinol 18: 2363–2377. doi: 10.1210/me.2003-0325.

19. **Ahmadian M, Suh JM, Hah N, Liddle C, Atkins AR, et al.** (2013) PPARγ signaling and metabolism: the good, the bad and the future. Nat Med 99: 557–566.

20. **Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, et al.** (2014) Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition. PLoS ONE 9: e99720. doi: 10.1371/journal.pone.0099720.

21. **Chen Z, Zhao X, Jiang X, Guo X, Lv Z, et al.** (2011) Association of PPARγ2 polymorphisms with carcass and meat quality traits in a Pietrain x Jinhua F2 population. Genet Mol Biol 34: 56–61.

22. **O'Gorman CW, Stanko RL, Keisler DH, Garcia MR** (2010) Effects of acute fasting and age on leptin and peroxisome proliferator-activated receptor gamma production relative to fat depot in immature and mature pigs. J Anim Physiol Anim Nutr 94: e266–e276. doi: 10.1111/j.1439-0396.2009.00968.x.

23. **Calero-Nieto FJ, Wood AD, Wilson NK, Kinston S, Landry J-R, et al.** (2010) Transcriptional regulation of Elf-1: locus-wide analysis reveals four distinct promoters, a tissue-specific enhancer, control by PU.1 and the importance of Elf-1 downregulation for erythroid maturation. Nucleic Acids Res 38: 6363–6374. doi: 10.1093/nar/gkq490.

24. **Chang P-Y, Miyamoto S** (2006) Nuclear Factor-κB Dimer Exchange Promotes a p21waf1/cip1 Superinduction Response in Human T Leukemic Cells. Mol Cancer Res 4: 101–112. doi: 10.1158/1541-7786.MCR-05-0259.

25. **Jin E, Liu J, Suehiro J, Yuan L, Okada Y, et al.** (2009) Differential roles for ETS, CREB, and EGR binding sites in mediating VEGF receptor 1 expression in vivo. Blood 114: 5557–5566. doi: 10.1182/blood-2009-05-220434.

26. **Baek K, Baek J-H** (2013) The transcription factors myeloid elf-1-like factor (MEF) and distal-less homeobox 5 (Dlx5) inversely regulate the differentiation of osteoblasts and adipocytes in bone marrow. Adipocyte 2: 50–54.

27. **Rudra D, deRoos P, Chaudhry A, Niec RE, Arvey A, et al.** (2012) Transcription factor Foxp3 and its protein partners form a complex regulatory network. Nat Immunol 13: 1010–1019. doi: 10.1038/ni.2402.

28. **Seale P, Bjork B, Yang W, Kajimura S, Chin S, et al.** (2008) PRDM16 controls a brown fat/skeletal muscle switch. Nature 454: 961–967. doi: 10.1038/nature07182.

29. **Berg F, Gustafson U, Andersson L** (2006) The Uncoupling Protein 1 Gene (UCP1) Is Disrupted in the Pig Lineage: A Genetic Explanation for Poor Thermoregulation in Piglets. PLoS Genet 2: e129. doi: 10.1371/journal.pgen.0020129.

30. **Chuikov S, Levi BP, Smith ML, Morrison SJ** (2010) Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. Nat Cell Biol 12: 999–1006. doi: 10.1038/ncb2101.

31. **Kim J, Cho I, Hong J, Choi Y, Kim H, et al.** (2008) Identification and characterization of new microRNAs from pig. Mamm Genome 19: 570–580. doi: 10.1007/s00335-008-9111-3.

32. **Kang HS, Okamoto K, Kim Y-S, Takeda Y, Bortner CD, et al.** (2011) Nuclear Orphan Receptor TAK1/TR4-Deficient Mice Are Protected Against Obesity-Linked Inflammation, Hepatic Steatosis, and Insulin Resistance. Diabetes 60: 177–188. doi: 10.2337/db10-0628.

33. **Collins LL, Lee Y-F, Heinlein CA, Liu N-C, Chen Y-T, et al.** (2004) Growth retardation and abnormal maternal behavior in mice lacking testicular orphan nuclear receptor 4. Proc Natl Acad Sci U S A 101: 15058–15063. doi: 10.1073/pnas.0405700101.

34. **Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, et al.** (2008) Many sequence variants affecting diversity of adult human height. Nat Genet 40: 609–615. doi: 10.1038/ng.122.

35. **Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, et al.** (2009) Meta-Analysis of Genome-Wide Scans for Human Adult Stature Identifies Novel Loci and Associations with Measures of Skeletal Frame Size. PLoS Genet 5: e1000445. doi: 10.1371/journal.pgen.1000445.

36. **Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, et al.** (2008) Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40: 575–583. doi: 10.1038/ng.121.

37. **Oettgen P, Akbarali Y, Boltax J, Best J, Kunsch C, et al.** (1996) Characterization of NERF, a novel transcription factor related to the Ets factor ELF-1. Mol Cell Biol 16: 5091–5106.

38. **Borowiec M, Liew CW, Thompson R, Boonyasrisawat W, Hu J, et al.** (2009) Mutations at the BLK locus linked to maturity onset diabetes of the young and β-cell dysfunction. Proc Natl Acad Sci 106: 14460–14465. doi: 10.1073/pnas.0906474106.

39. **Sutton P, Borgia J, Bonomi P, Plate J** (2013) Lyn, a Src family kinase, regulates activation of epidermal growth factor receptors in lung adenocarcinoma cells. Mol Cancer 12: 76.

40. **Abram CL, Courtneidge SA** (2000) Src Family Tyrosine Kinases and Growth Factor Signaling. Exp Cell Res 254: 1–13. doi: 10.1006/excr.1999.4732.

41. **Parsons JT, Parsons SJ** (1997) Src family protein tyrosine kinases: cooperating with growth factor and adhesion signaling pathways. Curr Opin Cell Biol 9: 187–192. doi: 10.1016/S0955-0674(97)80062-2.

42. **Udell CM, Samayawardhena LA, Kawakami Y, Kawakami T, Craig AWB** (2006) Fer and Fps/Fes Participate in a Lyn-dependent Pathway from FcεRI to Platelet-Endothelial Cell Adhesion Molecule 1 to Limit Mast Cell Activation. J Biol Chem 281: 20949–20957. doi: 10.1074/jbc.M604252200.

43. **Xiang P, Lo C, Argiropoulos B, Lai CB, Rouhi A, et al.** (2010) Identification of E74-like factor 1 (ELF1) as a transcriptional regulator of the Hox cofactor MEIS1. Exp Hematol 38: 798–808.e2. doi: 10.1016/j.exphem.2010.06.006.

44. **Economides KD, Zeltser L, Capecchi MR** (2003) Hoxb13 mutations cause overgrowth of caudal spinal cordand tail vertebrae. Dev Biol 256: 317–330. doi: 10.1016/S0012-1606(02)00137-9.

45. **Jung C, Kim R-S, Zhang H-J, Lee S-J, Jeng M-H** (2004) HOXB13 Induces Growth Suppression of Prostate Cancer Cells as a Repressor of Hormone-Activated Androgen Receptor Signaling. Cancer Res 64: 9185–9192. doi: 10.1158/0008-5472.CAN-04-1330.

46. **Norris JD, Chang C-Y, Wittmann BM, Kunder RS, Cui H, et al.** (2009) The Homeodomain Protein HOXB13 Regulates the Cellular Response to Androgens. Mol Cell 36: 405–416. doi: 10.1016/j.molcel.2009.10.020.

47. **Kamei Y, Miura S, Suzuki M, Kai Y, Mizukami J, et al.** (2004) Skeletal Muscle FOXO1 (FKHR) Transgenic Mice Have Less Skeletal Muscle Mass, Down-regulated Type I (Slow Twitch/Red Muscle) Fiber Genes, and Impaired Glycemic Control. J Biol Chem 279: 41114–41123. doi: 10.1074/jbc.M400674200.

48. **Armoni M, Harel C, Karni S, Chen H, Bar-Yoseph F, et al.** (2006) FOXO1 Represses Peroxisome Proliferator-activated Receptor-γ1 and -γ2 Gene Promoters in Primary Adipocytes: A Novel Paradigm to Increase the Insulin Sensitivity. J Biol Chem 281: 19881–19891. doi: 10.1074/jbc.M600320200.

49. **Huang TH, Zhu MJ, Li XY, Zhao SH** (2008) Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development. PLoS One 3: e3225.

50. **Leyme A, Bourd-Boittin K, Bonnier D, Falconer A, Arlot-Bonnemains Y, et al.** (2012) Identification of ILK as a new partner of the ADAM12 desintegrin and metalloprotease in cell adhesion and survival. Mol Biol Cell 23: 17 3461–3472. doi: 10.1091/mbc.E11-11-0918.

51. **Laumet G, Petitprez V, Sillaire A, Ayral A-M, Hansmannel F, et al.** (2010) A study of the association between the ADAM12 and SH3PXD2A (SH3MD1) genes and Alzheimer's disease. Neurosci Lett 468: 1–2. doi: 10.1016/j.neulet.2009.10.040.

52. **Bottini N, Vang T, Cucca F, Mustelin T** (2006) Role of PTPN22 in type 1 diabetes and other autoimmune diseases. Allelic Var Signal Elem Autoimmun 18: 207–213. doi: 10.1016/j.smim.2006.03.008.

53. **Wymann MP, Schneiter R** (2008) Lipid signalling in disease. Nat Rev Mol Cell Biol 9: 162–176. doi: 10.1038/nrm2335.

54. **Carrasco S, Mérida I** (2007) Diacylglycerol, when simplicity becomes complex. Trends Biochem Sci 32: 27–36. doi: 10.1016/j.tibs.2006.11.004.

55. **Berridge MJ, Lipp P, Bootman MD** (2000) The versatility and universality of calcium signalling. Nat Rev Mol Cell Biol 1: 11–21. doi: 10.1038/35036035.

56. **Hume JR, McAllister CE, Wilson SM** (2009) Caffeine inhibits InsP3 responses and capacitative calcium entry in canine pulmonary arterial smooth muscle cells. Vascul Pharmacol 50: 89–97. doi: 10.1016/j.vph.2008.11.001.

57. **Decrock E, De Bock M, Wang N, Gadicherla AK, Bol M, et al.** (2013) IP3, a small molecule with a powerful message. 12th Eur Symp Calcium 1833: 1772–1786. doi: 10.1016/j.bbamcr.2012.12.016.

58. **Bakthavatsalam D, Meijer HJG, Noegel AA, Govers F** (2006) Novel phosphatidylinositol phosphate kinases with a G-protein coupled receptor signature are shared by Dictyostelium and Phytophthora. Trends Microbiol 14: 378–382. doi: 10.1016/j.tim.2006.07.006.

59. **Lee TK, Lee C, Bischof R, Lambert GW, Clarke IJ, et al.** (2014) Stress-induced behavioral and metabolic adaptations lead to an obesity-prone phenotype in ewes with elevated cortisol responses. Psychoneuroendocrinology 47: 166–177. doi: 10.1016/j.psyneuen.2014.05.015.

60. **Allen DB** (1996) Growth suppression by glucocorticoid therapy. Endocrinol Metab Clin North Am 25: 699–717. doi: 10.1016/S0889-8529(05)70348-0.

61. **Serrano MP, Cámara L, Morales JI, Berrocoso JD, López Bote CJ, et al.** (2012) Effect of gender, housing density and the interaction on growth performance and carcass and meat quality of pigs slaughtered at 110 kg body weight. Span J Agric Res Vol 11 No 1 2013. Available: http://revistas.inia.es/index.php/sjar/article/view/2869.

62. **Mul D, Hughes IA** (2008) The use of GnRH agonists in precocious puberty. Eur J Endocrinol 159: S3–S8. doi: 10.1530/EJE-08-0814.

63. **Kantas D, Papatsiros V, Tassis P, Tzika E, Pearce MC, et al.** (2014) Effects of early vaccination with a gonadotropin releasing factor analog-diphtheria toxoid conjugate on boar taint and growth performance of male pigs. J Anim Sci 92: 2251–2258. doi: 10.2527/jas.2013-6924.

64. **Stevens A, Hanson D, Whatmore A, Destenaves B, Chatelain P, et al.** (2013) Human growth is associated with distinct patterns of gene expression in evolutionarily conserved networks. BMC Genomics 14: 547.

65. **Morgan MJ, Liu Z** (2011) Crosstalk of reactive oxygen species and NF-κB signaling. Cell Res 21: 103–115.

66. **Rubinek T, Rubinfeld H, Hadani M, Barkai G, Shimon I** (2005) Nitric oxide stimulates growth hormone secretion from human fetal pituitaries and cultured pituitary adenomas. Endocrine 28: 209–216. doi: 10.1385/ENDO:28: 2: 209.

67. **Dai Z, Wu Z, Yang Y, Wang J, Satterfield MC, et al.** (2013) Nitric oxide and energy metabolism in mammals. BioFactors 39: 383–391. doi: 10.1002/biof.1099.

68. **Dube A, Thai S, Gaspar J, Rudders S, Libermann TA, et al.** (2001) ELF-1 Is a Transcriptional Regulator of the Tie2 Gene During Vascular Development. Circ Res 88: 237–244. doi: 10.1161/01.RES.88.2.237.

69. **Cordani N, Pisa V, Pozzi L, Sciorati C, Clementi E** (2014) Nitric Oxide Controls Fat Deposition in Dystrophic Skeletal Muscle by Regulating Fibro-Adipogenic Precursor Differentiation. STEM CELLS 32: 874–885. doi: 10.1002/stem.1587.

70. **Serra X, Gil F, Pérez-Enciso M, Oliver M, Vázquez J, et al.** (1998) A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Livest Prod Sci 56: 215–223. doi: 10.1016/S0301-6226(98)00151-1.

71. **Owens FN, Dubeski P, Hanson CF** (1993) Factors that alter the growth and development of ruminants. J Anim Sci 71: 3138–3150.

72. **Gjerlaug-Enger E, Kongsro J, Ødegård J, Aass L, Vangen O** (2012) Genetic parameters between slaughter pig efficiency and growth rate of different body tissues estimated by computed tomography in live boars of Landrace and Duroc. animal 6: 9–18. doi: 10.1017/S1751731111001455.

73. **Pérez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, et al.** (2000) A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross. J Anim Sci 78: 2525–2531.

74. **Ramayo-Caldas Y, Castello A, Pena R, Alves E, Mercade A, et al.** (2010) Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics 11: 593.

75. **Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, et al.** (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. doi: 10.1371/journal.pone.0006524.

76. **Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al.** (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81: 559–575. doi: 10.1086/519795.

77. **Wigginton JE, Abecasis GR** (2005) PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. Bioinformatics 21: 3445–3447. doi: 10.1093/bioinformatics/bti529.

78. **Perez-Enciso M, Misztal I** (2011) Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 12: 202.

79. **Caraux G, Pinloche S** (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics 21: 1280–1281. doi: 10.1093/bioinformatics/bti141.

80. **Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM** (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10: 252–263. doi: 10.1038/nrg2538.

81. **Reverter A, Chan EKF** (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics 24: 2491–2497. doi: 10.1093/bioinformatics/btn482.

82. **Scardoni G, Petterlini M, Laudanna C** (2009) Analyzing biological network parameters with CentiScaPe. Bioinformatics 25: 2857–2859. doi: 10.1093/bioinformatics/btp517.

83. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al.** (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 13: 2498–2504. doi: 10.1101/gr.1239303.

84. **Benjamini Y, Hochberg Y** (1995) Controlling the false discovery rate: a pratical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57: 289–300.

85. **Storey JD, Tibshirani R** (2003) Statistical significance for genome-wide studies. Proc Natl Acad Sci 100: 9440–9445. doi: 10.1073/pnas.1530509100.

Expression-based GWAS identify variants, gene interactions and potential key regulators affecting the intramuscular content and fatty acid composition in porcine meat

Puig-Oliveras A, *et al*. (Manuscript in preparation)

Expression-based GWAS identify variants, gene interactions and potential key regulators affecting the intramuscular content and fatty acid composition in porcine meat

Anna Puig-Oliveras[1,2], *et al.* (Manuscript in preparation)

1 Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain, **2** Plant and Animal Genomics, Centre de Recerca en Agrigenòmica (CRAG), 08193 Bellaterra, Spain.

Email: anna.puig@cragenomica.es

Abstract

Background: The pork meat lean percentage and its quality determine the price paid for cuts by consumers. Meat with high intramuscular fat (IMF) composed of more monounsaturated fatty acid (MUFA) and, less polyunsaturated fatty acid (PUFA) and saturated fatty acid (SFA) content is considered of good quality being more tasty and juicy. Many genes determining fat content and composition in pork meat have been identified through quantitative trait loci (QTL) and genome-wide association analyses (GWAS); however, few genes containing causal mutations have been detected. The expression QTLs (eQTLs) are described to be heritable and to affect phenotypes. The aim of this work was to study the genetic basis that affects the abundance of transcripts in relation with lipid metabolism in swine muscle (*Longissimus dorsi*) in an Iberian × Landrace (IBMAP) backcross (BC1_LD).

Results: In this study we performed a genome-wide association with the PorcineSNP60 BeadChip genotype information and the mRNA expression levels (eGWAS) of 45 lipid-related genes measured by Real-Time PCR in 114 BC1_LD animals. The eGWAS identified 241 eSNPs located in 18 chromosomal regions on SSC1, SSC2, SSC3, SSC6, SSC8, SSC9, SSC10, SSC11, and SSC13 and associated with 11 genes *ACSM5, CROT, FABP3, FOS, HIF1AN, IGF2, MGLL, NCOA1, PIK3R1, PLA2G12A,* and *PPARA.* Three eQTLs for *IGF2, ACSM5* and *MGLL* were identified showing *cis*-acting effects, whereas 16 eQTLs had *trans* regulatory effects on gene expression traits for *ACSM5, CROT, FABP3, FOS, HIF1AN, MGLL, NCOA1, PIK3R1, PLA2G12A,* and *PPARA* genes. A total of 46% of the 241 eSNPs identified were located within a gene. Strong candidate genes regulating *ACSM5, FOS, PPARA, PIK3R1, PLA2G12A* and *HIF1AN* gene expression were evidenced with the eGWAS. Furthermore, potential regulators co-localizing within fatness and growth related QTLs previously identified in the IBMAP cross (i.e. *ARHGAP6, IGF2, MC2R, MGLL, NR3C1*) were also evidenced. In addition, the NR3C1 transcription factor was identified as a strong candidate gene regulating the 45 lipid metabolism related genes analyzed.

Conclusions: The obtained results highlighted genes and pathways that may be key in determining the IMF content and fatty acid (FA) composition in the IBMAP backcross and increased our knowledge in the functional regulatory mechanisms implicated in these complex traits.

Keywords: eQTL, *Longissimus dorsi*, gene expression, lipid metabolism

Introduction

Pork meat cuts and their derived products are paid according to the lean percentage in pork carcasses and the meat quality since they determine a better acceptance for consumers (Schwab *et al.,* 2006). A high amount of backfat content is a less desirable trait; meanwhile, meat with high IMF is considered to have better taste conferring juiciness to the meat. Besides, FA composition of IMF affects the meat nutritional and sensory quality parameters. MUFA confers more oxidative stability than PUFA improving meat taste and colour (Wood *et al.,* 2008). Furthermore, PUFA decreases the risk of suffering cardiovascular diseases being healthier than SFA (Michas *et al.,* 2014). Thereby, there is a consumer requirement for porcine meat with high IMF with a balanced FA composition.

Selected breeds for the pig industry as Landrace have an efficient meat production with a rapid growth and leaner carcass; however, the resulting meat has low IMF and high PUFA content (Estévez *et al.,* 2003). In contrast, other breeds as Iberian have a slow growth and are used to produce high quality and tasty meat, especially for dry-cured products, showing more IMF with higher MUFA and SFA content. The differences in the genetic background of these breeds determine the IMF and its FA composition affecting the pork meat quality (Serra *et al.,* 1998).

The experimental IBMAP BC1_LD, 25% Iberian and 75% Landrace population, was generated to study the genetic mechanisms that determine relevant production traits for the porcine meat industry. Strategies such as QTL and GWAS have been useful to highlight many genes determining IMF content and composition in the IBMAP population (Pérez-Enciso *et al.,* 2000; Óvilo *et al.,* 2002; Clop *et al.,* 2003; Mercadé *et*

al., 2005a; Muñoz *et al.,* 2011; Fernández *et al.,* 2012; Ramayo-Caldas *et al.,* 2012; Muñoz *et al.,* 2013) resulting in the identification of several candidates genes (Estellé *et al.,* 2005; Mercadé *et al.,* 2005b; Estellé *et al.* 2006, Mercadé *et al.,* 2006; Corominas *et al.,* 2012, Corominas *et al.,* 2013; Revilla *et al.,* 2014). Despite this, the underlying physio-genetic complex mechanisms of the IMF deposition and its FA composition have not been clarified. The difficulty in the detection of QTLs for complex traits may be influenced by the pleiotropic nature of these traits and the use of multiple tests correction methods (Rao 2008).

The detection of eQTLs has recently been proposed as a good strategy to deepen in the study of the genetic architecture of complex traits (Schadt *et al.,* 2003; Gilad *et al.,* 2008). This technique allows the identification of genetic variants associated with gene transcription levels which may be determining the phenotypic differences of complex traits. With the aim to better understand the mechanisms affecting the IMF content and FA composition, we performed eGWAS of 45 strong candidate genes identified in previous studies of our group in 114 BC1_LD animals.

Results and Discussion

Selection of lipid-related metabolism genes in muscle

In previous studies of our group, strong candidate genes affecting the IMF content and FA composition of the *Longissimus dorsi* muscle of the IBMAP BC1_LD were identified by using GWAS, RNA-Seq and co-association network approaches (Ramayo-Caldas *et al.,* 2012; Puig-Oliveras *et al.,* 2014a; Puig-Oliveras *et al.,* 2014b; Ramayo-Caldas *et al.,* 2014). In the present study, a list of 45 genes functionally related with lipid metabolism was selected, prioritizing candidate genes for FA composition.

We included (1) candidate genes differentially expressed (*ACAA2, AQP7, ALB, ANGPT1, ATF3, MLXIPL, FOS, HIF1AN, PIK3R1, PLIN5, PPARG, SCD, SLC2A4*) in the *Longissimus dorsi* muscle of two phenotypically extreme groups of animals for intramuscular FA composition from the IBMAP cross and their potential regulators (*NFKB1, PPARGC1A*) (Puig-Oliveras *et al.*, 2014a). (2) Candidate functional and positional genes *(FABP5, PIK3R1, PLA2G12A, PPAP2A*) identified in a GWAS study for intramuscular FA

composition in the same animal material (Ramayo-Caldas *et al.,* 2012); and (3) genes related to lipid metabolism identified in gene co-association networks for FA composition (*ACSM5, ANGPT1, FABP3, FABP5, MGLL, NCOA2, PEX2, PPARG, SETD7*; Ramayo-Caldas *et al.,* 2014) and fatness and growth related traits (*ALB, CREG1, ELF1, FABP5, MGLL, PPARG*; Puig-Oliveras *et al.,* 2014b). Finally, in order to complete the set of genes, we chose genes which have been described in the literature to play different roles in muscle lipid metabolism such as transcriptional factors, cofactors and nuclear receptors (*ETS1, LPIN1, NR1H3, NCOA1, NCOA6, PPARA, PPARD, PRKAA1, RXRG, SP1, SREBF1*) (http://www.bioguo.org/AnimalTFDB/; Zhang *et al.* 2012), enzymes (*ACSS1, ACSS2, CPT1B, CROT, DGAT1, DGAT2*, *PDHX*) as well as the *IGF2* gene, which was described as the causal factor of the imprinted QTL for muscle growth and fat deposition in a Meishan × Large White intercross (Van Laere *et al.,* 2003). Moreover, polymorphisms in *IGF2* gene have been significantly associated to muscle FA composition in swine (Hong *et al.,* 2015).

Gene expression profiling

In the present study, 114 backcrossed (BC1_LD) animals generated from the experimental IBMAP population and showing a wide range of IMF content and FA composition values were used for RNA extraction and Real-Time PCR to perform the eGWAS (Table S1).

Sex significant effect (p-value≤0.05) between gene expression levels was detected in 20 out of the 45 analyzed genes (44%): *ACSS1, ACSS2, ATF3, CREG1, DGAT2, ETS1, FABP5, HIF1AN, IGF2, NCOA2, NCOA6, PLA2G12A, PPARA, PPARG, PPARGC1A, PRKAA1, PEX2, SCD, SP1,* and *SREBF1* (Figure S1). Sexually dimorphic gene expression in genes involved in lipid metabolism has already been described in muscle as in other tissues such as liver (Liu *et al.,* 2010; Zhang *et al.,* 2011). Interestingly, some of the sex-biased genes identified are key regulators of lipid metabolism in muscle, such as *PPARA, PPARG, PPARGC1A* and *SREBF1*. Notice that within the sex-biased genes, several lipogenic genes were more expressed in females (*DGAT2, NCOA2, NCOA6, PPARG, PRKAA1, SCD, SP1* and *SREBF1*), whereas more lipolytic genes were over-expressed in males (*ATF3, PPARA* and *PPARGC1A*). Consistent with these results, *PPARA* and *SREBF1*

presented sex-biased gene expression in human skeletal muscle (Fu *et al.,* 2009). Furthermore, *PPARGC1A* and *PPARA* gene expression in muscle are affected under diet supplementation with 17β-estradiol (E_2) (Fu *et al.,* 2009). In addition, the mRNA expression of *SCD* and *PPARG* genes has been observed to be lower in male than in female human muscle cell cultures (Rune *et al.,* 2009).

To assess the relationship between muscle gene expression levels and phenotypes, a hierarchical cluster analysis of the correlation values among the gene expression levels (RQ) of the 45 genes and the fatty acid content in muscle was performed. The hierarchical cluster analysis showed that genes related mainly with lipogenic pathways (*DGAT2, PPARG, SCD, MGLL, NCOA1, NCOA2, NCOA6, PRKAA1, SP1, SREBF1*) clustered together; whereas a second cluster was found for genes mainly related with lipolytic pathways (*ATF3, CPT1B, PPARA, PPARD, PPARGC1A*) (Figure 1). Genes clustering within the lipogenic-related cluster showed in general positive correlations with palmitoleic (C16:1(n-7)) and octadecenoic (C18:1(n-7)) FAs; while the lipolytic-related group showed in general a positive correlation with PUFAs and in special linoleic (C18:2(n-6)) FA. In general, these results are in agreement with a previous muscle RNA-Seq transcriptome study (Puig-Oliveras *et al.,* 2014a) of animals extreme for intramuscular FA composition performed by our group where a higher expression of genes related with lipogenic pathways was observed in the muscle of animals with high MUFA and SFA content in muscle.

The highest correlations (p-value<1.00×10^{-16}) were observed for *PPARG* and *SCD* (r=0.78), *DGAT2* and *PPARG* (0.85), *SCD* and *DGAT2* (r=0.77), *FABP3* and *PLIN5* (r=0.83), *FABP3* and *AQP7* (r=0.80), *ACAA2* and *FABP3* (r=0.78) and *ELF1* and *PPAP2A* (r=0.75). Interestingly, the strong correlation of mRNA expression of *PPARG* and *SCD* (r=0.78; p-value<1.00×10^{-16}) suggests a transcriptional regulation of *SCD* by the PPARG nuclear factor. Supporting our results, recent results obtained in dairy goat have showed that PPARG contributes to the regulation of *SCD* in mammary epithelial cells (Shi *et al.,* 2013).

Here, *PPARG, DGAT2* and *SCD*, which are involved in the triacylglycerol synthesis, were highly correlated. Similarly, *FABP3*, *AQP7* and *PLIN5,* which encode for proteins

responsible of lipid and glucose transport, were also highly correlated. Noteworthy, these three genes were over-expressed in animals with higher MUFA and SFA content in muscle when compared with animals having more PUFA in a RNA-Seq study (Puig-Oliveras *et al.,* 2014a).

eQTL identification

An eGWAS was performed using a total of 40,586 SNPs and the mRNA expression values of the 45 lipid-related genes of the 114 BC1_LD animals. A total of 241 eSNPs in 18 chromosomal regions located in SSC1, SSC2, SSC3, SSC6, SSC8, SSC9, SSC10, SSC11, and SSC13 were identified for a total of 11 genes: *ACSM5, CROT, FABP3, FOS, HIF1AN, PIK3R1, PLA2G12A, MGLL, IGF2, NCOA1* and *PPARA* (FDR<0.05; Table 1). Five genes (*ACSM5, IGF2, MGLL, PLA2G12A* and *PPARA*) presented more than one associated eQTL (Table 1). Three out of 18 eQTLs were identified as *cis*-acting for *ACSM5, IGF2* and *MGLL* gene expression (Figure 2), suggesting the presence of a mutation in the same gene directly affecting its expression, whereas 16 eQTLs had *trans* regulatory effects. The majority of eQTLs (8 of 18) were located in SSC2 and SSC6 (Table 1).

From the associated eSNPs (n=241), 215 eSNPs were successfully annotated with the Variant Effect Predictor of Ensembl (VEP; Sscrofa 10.2 annotation release 80) of which 54% (117 eSNPs) were located in intergenic regions. The remaining 46% (98) eSNPs mapped within 76 genes: 69 (32%) in introns, 10 in the 5' flanking region, 13 in the 3' flanking region, three in the 3'UTR region, two in the coding region of a gene determining synonymous mutations and one missense mutation. Twelve out of the 98 intra-genic eSNPs were located within a gene exerting a lipid metabolism function (*ABCA3, ACSM3, LRP5, LHFPL4, BRPF1, HRH1, PPARG, ACAD9, COPG1, MGLL, ACAD11, ARHGAP26*) based on Genecards [http://www.genecards.org] and Mouse Genome Informatics database [http://www.informatics.jax.org] (Table S2). Of these twelve genes related to the lipid metabolism, only six of them were described to be expressed in muscle (*ACSM3, PPARG, ACAD9, MGLL, ACAD11*, and *ARHGAP26*). Only one of the intragenic eSNPs mapping within genes with a lipid-related function was identified in a *trans* eQTL: the *ARHGAP26* gene located in a *trans*-eQTL for *PPARA* gene. This gene is activated via lipid interaction, however its role has not been well defined (Erlmann *et*

al., 2009). The other eSNPs located within a gene having a lipid-related function were inside *cis*-eQTLs, where the strong candidate gene was the analyzed gene.

cis-eQTL identified

The *insulin-like growth factor 2* (*IGF2*) gene is not mapped in the current *Sscrofa10.2* assembly (Rubin *et al.,* 2012). However, it has been located by linkage map in the telomeric end of the p arm of SSC2 (Fontanesi *et al.,* 2010). An intronic *IGF2* mutation (*IGF2*-intron3-G3072A) was described by Van Laere *et al.* (2003) to have a major effect on muscle growth in pig. Although this mutation was segregating only in a small family on the IBMAP F_2 population (Estellé *et al.,* 2005), the *cis*-eQTL for *IGF2* suggests that the intronic mutation identified by Van Laere *et al.* (2003) or another variant within the *IGF2* gene may be segregating in the IBMAP BC1_LD, having an important effect on the analyzed traits. Further analysis will be conducted to validate our hypothesis. Concerning the *monoglyceride lipase* (*MGLL*) eGWAS results, one of the annotated *cis*-eSNP (ASGA0103932) mapped within an intronic region of *MGLL* gene (Table S2). However, this SNP was not the most significantly associated SNP (ASGA0103932; p-value=2.34×10^{-8}), suggesting the presence of other SNPs within or near this gene as causative mutation with effects in the *MGLL* gene expression levels. The most significant *cis*-eSNP for *MGLL* (ASGA0093606; p-value=2.20×10^{-9}) was located less than 0.69Mb downstream of the *MGLL* gene. In addition, an eSNP in the *MGLL cis*-eQTL was annotated in the upstream region of *PPARG* gene (Table S2). *PPARG* and *MGLL* genes were reported to be co-associated in a previous study of our group for growth and fatness traits, where *PPARG* was described as a major regulator (Puig-Oliveras *et al.,* 2014b). Besides, the literature-based analysis with Genomatix [http://www.genomatix.de/] also identified an interaction between these two genes (Figure S2), and a chromatin immunoprecipitation (ChiP) experiment performed in epithelial cultured cells revealed PPARG binding sites in the distal *MGLL* promoter (Harmon *et al.,* 2010). Though, it is likely that this region initially considered as a single *cis*-acting eQTL comprises in fact two eQTLs. In this regard, the *MGLL* gene expression would be affected by a variant present in the same gene (*MGLL*) and also by *PPARG*. Reinforcing this hypothesis, we observed that it was the largest *cis* interval (comprising approximately 54.6Mb) in comparison with the two other *cis* intervals (spanning a

maximum length of 11.5Mb) identified for the *IGF2* and *ACSM5* genes (Table 1). The correlation between *PPARG* and *MGLL* gene expression in muscle was moderate (r=0.48; p-value=8.28×10^{-8}). No association was found between the SNP within *PPARG* (ISU10000701) and the *PPARG* gene expression, suggesting than other polymorphism in this region is responsible for the variation in *MGLL* gene expression. Remarkably, the significant associated eSNP inside the *PPARG* gene (ISU10000701) was one of the three main co-associated SNPs in the network identified in Puig-Oliveras *et al.* (2014b) involved in the determination of growth and fatness traits.

Three eSNPs (ASGA0090088, ASGA0105223 and SIRI0001454) in complete linkage disequilibrium (D'=1) were the most significant in the *cis*-eQTL (p-value=7.12×10^{-14}) associated with the mRNA levels of the *acyl-coA synthetase medium-chain family member 5* (*ACSM5*) (Table 1). The ASGA0090088 marker was the closest *cis*-eSNP mapping at approximately 798kb of the upstream region of the *ACSM5* gene (Table S2).

trans-eQTL identified

The *ACSM5* gene expression was associated in *trans* with two chromosomal regions in SSC3 and SSC10. The most associated SNP of SSC10 was an intronic polymorphism (ASGA0090778; p-value=2.55×10^{-8}) in the *COG7* gene (Table S2). Although the *COG7* gene has not been described to have a direct relationship with the lipid metabolism function, it is reported that members of the conserved *component of oligomeric golgi complex* (COG) are involved in intra-Golgi trafficking and glycosylation of proteins and lipids (Smith & Lupashin 2008). Other lipid-related genes were identified within this *trans*-eQTL in SSC10 such as the *NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa* (*NDUFAB1*) gene which plays a role in FA biosynthesis and the *golgi-associated, gamma adaptin ear containing, ARF binding protein 2* (*GGA2*) gene, which is a component of the clathrin coats involved in the lipid membrane exchange (Hung *et al.,* 2012) (Table S3). Within the second *trans*-eQTL for *ACSM5* at 100.35Mb of SSC3, two genes that may affect the lipid metabolism were identified: *protein kinase c, epsilon* (*PRKCE*) and *calmodulin 1 (Phosphorylase kinase, delta)* (*CALM1*) (Table S3). *PRKCE* knockdown led to an increase of FA esterification in hepatocytes and *CALM1*

gene expression (Raddatz *et al.,* 2011). CALM1 is a subunit of the phosphorylase kinase enzyme, a rate-limiting enzyme for glycogenolysis (Mormeneo *et al.,* 2012).

For *FBJ murine osteosarcoma viral oncogene homolog* (*FOS*) gene expression, three interesting genes in a SSC11 *trans*-eQTL were identified: *StAR-related lipid transfer (START) domain containing 13* (*STARD13*), *spastic paraplegia 20 (Troyer syndrome)* (*SPG20*), and *arachidonate 5-lipoxygenase-activating protein* (*ALOX5AP*) (Table S3). The StAR gene family encode for globular proteins that form cavities where lipids and lipid hormones bind to be exchanged between biological membranes (Thorsell *et al.,* 2011). Supporting STARD13 as a regulator of *FOS* gene expression, in transgenic mice with pancreas *STARD13* ablation there was no detectable mRNA expression of *FOS* gene (Petzold *et al.,* 2013). *SPG20* may be involved in increased lipid droplet and alterations in perilipin levels (Renvoisé *et al.,* 2012). Finally, ALOX5AP is required for leukotriene (an arachidonic metabolite) synthesis implied in inflammatory responses.

Two regions in SSC2 and SSC6 were associated in *trans* with the *peroxisome proliferator-activated receptor alpha* (*PPARA*) gene expression. In SSC2, several genes related with lipid metabolism were identified: *Rho GTPase activating protein 26* (*ARHGAP26*)*, fibroblast growth factor 1 (Acidic)* (*FGF1*) and *nuclear receptor subfamily 3, group c, member 1 (glucocorticoid receptor)* (*NR3C1*) (Table S3). The *ARHGAP26* is activated via lipid interaction (Erlmann *et al.,* 2009) and may play a role in adipogenesis, as well as the transcription factor *NR3C1*, which maps at approximately 0.5Mb of *ARHGAP26* gene. By using the Genomatix interface [http://www.genomatrix.de/], we observed that NR3C1 interacts with *NFKB*, *CREB* and *NCOA* genes, and may be affected by the corticotropin-releasing hormone (CRH) and insulin (INS) (Figure 3). Accordingly, the CRH was identified in a network analysis for fatness and growth traits (Puig-Oliveras *et al.,* 2014b) and the insulin signalling pathway in a RNA-Seq study comparing animals extreme for their intramuscular FA composition (Puig-Oliveras *et al.,* 2014a), both studies performed with the IBMAP animal material. The *FGF1* gene has been identified as differentially expressed in animals phenotypically extreme for FA composition in muscle (Puig-Oliveras *et al.,* 2014a). Noteworthy, FGF1 is involved in preadipocyte differentiation and has been suggested to be acting on the PPARG system, however the mechanisms remain unclear

(Hutley *et al.,* 2004). In accordance with this hypothesis, it is described that FGF1 can be trans-located to the nucleus exerting a growth regulatory activity (Zhen *et al.,* 2012). Hence, it would be important to study the FGF1-PPARA relationship. In the SSC6 *PPARA* trans-eQTL mapped the *palmitoyl-protein thioesterase 1* (*PPT1*) gene, which is involved in the catabolism of lipid modified proteins (Calero *et al.,* 2003), the *metallophosphoesterase 1* (*MPPE1*) which acts in lipid remodelling of GPI-anchor proteins, the *inositol(myo)-1(or 4)-monophosphatase 2* (*IMPA2*) which plays an important role in phosphatidylinositol signalling, and the *cell death-inducing DFFA-Like effector A* (*CIDEA*) involved in lipolysis and thermogenesis. Interestingly, *CIDEA*-null mice showed a decreased *PPARA, PPARG* and *SREBF1* gene expression and a decreased *de novo* FA synthesis in liver (Zhou *et al.,* 2012). Besides, the *CIDEA*-*PPARA* interaction identified in the eQTL analysis was also captured by Genomatix [http://www.genomatrix.de/] literature-based analysis (Figure S2). In addition, two melanocortin receptor genes (*MC2R* and *MC5R*) mapped in the *PPARA* SSC6 *trans*-eQTL. In adipocyte cells of *MC2R* knockdown mice alterations in fatty acid composition were observed: a reduction in the C16:1/C16:0 and C18:1/C18:0 ratios and an increase in the arachidonic acid content (Betz *et al.,* 2012). *MC2R* is specifically activated by the adrenocorticotropic hormone (ACTH) (Agulleiro *et al.,* 2013). On the other hand, MC5R has been linked or associated with the obesity phenotypes such as body mass index and fat mass in humans (Chagnon *et al.,* 1997).

The *phosphoinositide-3-kinase, regulatory subunit 1 alpha* (*PIK3R1*) eGWAS revealed a *trans*-eQTL on SSC6 at 145.60Mb where several interesting genes involved in lipid metabolism were mapped: *low density lipoprotein receptor-related protein 8, apolipoprotein e receptor* (*LRP8*), *24-dehydrocholesterol reductase* (*DHCR24*), *dab reelin signal transducer homolog 1* (*DAB1*), *protein kinase, AMP-activated, alpha 2 catalytic subunit* (*PRKAA2*), *phosphatidic acid phosphatase type 2B* (*PPAP2B*), and *proprotein convertase subtilisin/kexin type 9* (*PCSK9*) (Table S3). The *DHCR24* gene regulates the production of cholesterol (Mirza *et al.,* 2006), the *PPAP2B* is involved in the diacylglycerol synthesis for *de novo* lipogenesis of glycerolipids, and the PCSK9 regulates the plasma cholesterol homeostasis and binds to low and very low density lipoprotein receptors (LDLR and VLDLR) and apolipoprotein receptors (LPR1 and LPR8)

for their degradation. *DAB1* is involved in intracellular signalling pathways, including the PI3K signalling inducing its phosphorylation (Herz *et al.,* 2009). However, this gene was detected at very low expression levels in the muscle tissue of the RNA-Seq animals (Puig-Oliveras *et al.,* 2014a). Interestingly, the *PRKAA2* gene identified within the SSC6 *trans*-eQTL for *PIK3R1*, is a catalytic subunit of the AMP-activated protein kinase (AMPK) which is in charge of regulating lipid synthesis by phosphorylating lipid metabolic enzymes such as *ACACA, ACACB, ACC, GYS1, HMGCR, HSL* and *LIPE* to inactivate them. Therefore, it acts regulating key enzymes of fatty acid uptake, esterification, lipolysis and oxidation (O'Neill *et al.,* 2013). In the same direction, *PRKAA2* knockdown affects Akt activation. Therefore, PIK3R1 and PRKAA2, are both associated with the PI3K-Akt signalling pathway. Supporting these results, recent published studies suggested that AMPK activates Akt via regulating PI3K (Tao *et al.,* 2010). These results highlighted the *PRKAA2* as strong candidate gene to explain the variation in the mRNA levels of *PIK3R1*.

Four chromosomal regions (two regions in SSC6, one region in SSC8, and one in SSC9) where associated in *trans* with the *phospholipase A2, group XIIA* (*PLA2G12A*) gene expression. Specifically, on the SSC8 eQTL at approximately 128 to 136Mb were mapped the *microsomal triglyceride transfer protein* (*MTTP*)*, the hematopoietic prostaglandin D synthase* (*HPDGS*)*,* and the *alcohol dehydrogenases 4, 5 and 7* (*ADH7, ADH4,* and *ADH5*) genes involved in lipid metabolism (Table S3). The *MTTP* exhibiting a lipid transfer activity has been associated with C16:0, palmitoleic (C16:1(n-7)) and C18:1(n-9) FAs (Estellé *et al.,* 2009; Ramayo-Caldas *et al.,* 2012). The alcohol dehydrogenases are responsible for catalyzing ethanol and acetic acid to acetyl-CoA, needed for FA synthesis (Montooth *et al.,* 2006). The *HPGDS* gene is identified to catalyse the conversion of prostaglandins (lipid derived signalling molecules) to prostanoids involved in immune response (Virtue *et al.,* 2015). The *ADH5* was the most expressed gene in pig muscle (Puig-Oliveras *et al.,* 2014a); meanwhile, the other ADH members (ADH7 and ADH4) jointly with, *MTTP* and *HPGDS* genes showed very low expression levels in muscle (Puig-Oliveras *et al.,* 2014a) what is in concordance with what is reported in human muscle (Kapushesky *et al.,* 2010). Two genes were identified within the *trans*-eQTL at 79.9Mb position on SSC6: *mitochondrial trans-2-*

enoyl-coA reductase (*MECR*) and *sestrin 2* (*SESN2*) (Table S3). *MECR* is involved in FA synthesis and transcription modulation of PPARs (Parl *et al.,* 2013); and *SESN2* may contribute to the inhibition of *the nuclear receptor subfamily 1, group H, member 3 mediated* (*NR1H3*-mediated or *LXRα*) hepatic lipogenesis (Jin *et al.,* 2013). Only one lipid related gene was identified within the second *trans*-eQTL at 9.7Mb the *WW domain containing oxidoreductase* (*WWOX*) gene which has recently been described to play a role in cholesterol homeostasis and triglyceride biosynthesis (Iatan *et al.,* 2014).

Noteworthy, an eQTL on SSC9 at around 117Mb was affecting the expression of genes *PLA2G12A* and the *hypoxia inducible factor 1, alpha subunit inhibitor* (*HIF1AN*) (Table 1). Their mRNA expression was highly correlated (r=0.60; p-value=1.98×10^{-12}), suggesting a common element regulating their transcriptional levels. The three most significant SNPs (H3GA0028012, ASGA0044215, ALGA0117195; p-value=4.48×10^{-6}) within this eQTL were in complete linkage disequilibrium (D'=1) (Table 1). In this region was mapped the *PIK3CG* which encodes a *phosphatidylinositol-4,5-bisphosphate 3-kinase* which is known to participate in different functions including the insulin signalling pathway and the lipid metabolism (Kobayashi *et al.,* 2011), and the *DLD* gene which encodes a *dihydrolipoamide dehydrogenase* involved in acetyl-coA biosynthesis (Table S3). The *PIK3CG* gene, unlike other PI3K family members, is activated by interaction with G-protein-coupled receptors and silencing this gene the PI3K-Akt signalling pathway is inhibited (Semba *et al.,* 2002). In accordance, studies in cell lines have suggested that *HIF1AN* gene expression is repressed by a mechanism involving PI3K signalling (Datta *et al.,* 2004). Altogether, these results suggest that the PRKAA2 may regulate the class I PI3K regulatory subunit 1 (PIK3R1), which is able to form a heterodimer with the PIK3CG catalytic subunit, activate Akt pathway and inhibit *HIF1AN* gene expression.

For *IGF2* gene expression, apart from the SSC2 *cis*-eQTL, a *trans* eQTL located approximately at 162Mb of the same chromosome was identified and the *sirtuin 3* (*SIRT3*) gene mapped in this region (Table S3). It has been described that *SIRT3* knockout mice exhibited decreased oxygen consumption and increased oxidative stress in skeletal muscle (Jing *et al.,* 2011). Moreover, *SIRT3* knockout mice showed a down-regulation of the Akt phosphorylation (Jing *et al.,* 2011) (Figure S3). Members of

the *SIRT* gene family have been described to be relevant genes controlling lipolysis and promoting fat mobilization in white adipose tissue (Schug & Li 2011) and *SIRT1* member has been identified as one of the most central genes in a liver co-expression network for intramuscular FA composition in the IBMAP animal material (Ramayo-Caldas *et al.,* 2014). Recently, the *SIRT3* member, which can be activated by the AMPK protein, has been suggested to play a major role in obesity-related diseases (Newsom *et al.,* 2013).

Finally, within the *trans*-eQTLs identified for *CROT, FABP3*, *MGLL,* and *NCOA1* gene expression, we could not detect any strong candidate gene exerting a known lipid metabolism function.

Functional network analysis of genes mapping in eQTLs

For *trans*-eQTLs all the genes located within a ±1Mb interval were selected for gene annotation. Conversely, for *cis*-eQTLs only the studied candidate gene was considered (*ACSM5*, *IGF2*, and *MGLL*) for further analyses. In the 18 eQTLs, a total of 292 protein-coding genes, 13 miRNA, one miscRNA, six pseudogenes, one rRNA, eleven snoRNAs and four snRNAs were annotated. From the 292 protein-coding genes with Ensembl Gene ID, 256 had at least one human orthologous gene and were submitted to IPA to perform a functional categorization (Table S4).

The main networks identified with IPA analysis were: (i) energy production, small molecule biochemistry, and drug metabolism (score=44); (ii) organismal injury and abnormalities, cancer, and hematological disease (score=42); (iii) and connective tissue disorders, inflammatory disease, skeletal and muscular disorders (score=37) (Table S4).

Focusing on the first network comprising the energy production, small molecule biochemistry and drug metabolism functions, we identified the serine/threonine kinase effector (Akt) complex as central in the network (Figure S3). Remarkably, the Akt complex, which is involved in glucose transport and lipogenesis, was also identified in the muscle transcriptome study between 12 BC1_LD animals extreme for intramuscular FA composition (Puig Oliveras *et al.,* 2014a). In agreement with this results, several genes (*PIK3CG, PPAP2B, PRKAA2, PTPN2* and *SIRT3*) identified as

potential regulators within the eQTLs of the 45 lipid-related genes, are strongly related to the Akt pathway.

Identification of master regulators

A total of 298 genes including (1) the 253 genes annotated in the *trans*-eQTL intervals, and (2) the 45 studied genes of the present study were analyzed with iRegulon cytoscape plugin (Janky *et al.,* 2014). We observed that the *EP300* gene was the most enriched transcription factor motif (enrichment score threshold for the motif, NES=4.737). Noteworthy, the *EP300* gene was identified as a key regulator of FA composition and IMF traits in the same material using a gene co-association network (Ramayo-Caldas *et al.,* 2014). On the other hand, five of the 45 studied genes were identified as strong regulators of the 45 genes with iRegulon: *NR1H3* (NES=5.348; 10 target genes), *MLXIPL* (NES=4.988; 29 target genes), *PPARA* (NES=4.785; 32 target genes), *NFKB1* (NES=3.855; 6 target genes) and *PPARG* (NES=3.536; 10 target genes). The PPARA motif was present on the highest number of target genes (32 out of 45 genes). Interestingly, *NR3C1* gene identified within the *PPARA trans*-eQTL on SSC2 was also identified with iRegulon among the regulators for the 45 analyzed genes and MatInspector (Genomatix software) predicted a binding-site for this transcription factor in the promoter of *PPARA*. The *NR3C1* gene may play a negative role in adipogenesis, regulating the lipolytic and antilipogenic gene expression. In human studies, polymorphisms in the *NR3C1* gene have been suggested to contribute to obesity (Dobson *et al.,* 2001). Thus we hypothesize that the *NR3C1* may be a master regulator of lipid metabolism through the regulation of *PPARA* gene expression. However, further studies are necessary to corroborate this hypothesis.

QTL and eQTL co-localization

Finally, we looked for the overlapping of the 241 eSNPs with the QTLs described in the Pig QTLdb [http://www.animalgenome.org/cgi-bin/QTLdb/SS/index; Release 27, Apr 27, 2015]. A total of 234 eSNPs (97%) co-localized in 132 QTLs for fatness traits and 157 eSNPs (65%) within 10 different QTLs for FA composition (Table S5), confirming a high co-localization of eQTLs and fat-related QTLs. The large number of QTLs described for these traits, covering a big extension of the porcine genome, provides evidences for a complex genetic pleiotropic regulation basis of these traits.

Remarkably, five genes identified within the eQTLs (i.e. *ARHGAP6, IGF2, MC2R, MGLL, NR3C1*) overlapped with described QTLs in the IBMAP population. For instance, the *IGF2* gene for which a *cis*-eQTL was detected, was identified within a confidence interval of one of the epistatic regions affecting muscle fiber traits (Estellé *et al.,* 2008). The *MGLL* gene, showing a *cis*-acting eQTL, was located close to a GWAS interval affecting FA composition traits: palmitic (C16:0) and oleic, (C18:1(n-9)) FAs, polyunsaturated/saturated ratio (PUFA/SFA ratio), SFA, and unsaturated index (UI) (Ramayo-Caldas *et al.,* 2012). Furthermore, *MGLL* maps within a QTL interval affecting growth (Fernández *et al.,* 2012). The *MC2R* gene, identified in a *PPARA* trans-eQTL, maps within an IBMAP QTL region for IMF content and backfat thickness (Óvilo *et al.,* 2000). Moreover, *ARHGAP6* and *NR3C1* genes also identified in a *trans*-eQTL for *PPARA* gene expression are located within a QTL for growth traits in the IBMAP population (Fernández *et al.,* 2012).

Conclusions

To gain insight into the genetic control of lipid metabolism-related traits, potential candidate genes and variants regulating the transcriptional level of 45 lipid related-genes were identified using eQTL mapping and functional analyses. This study has identified several genomic regions containing candidate genes that may regulate gene expression levels of relevant lipid-metabolism related genes. Combined assessment of the obtained results from eGWAS of lipid-related genes and GWAS may provide

complementary information of genes and variants determining the IMF and FA composition traits. Therefore, the present results identify potential key genes and variants affecting pork meat quality. However, more efforts should be made to validate our results, for instance, the implication of the *NR3C1* gene as a major regulator in muscle FA metabolism.

Methods

Animal samples and phenotypes

The IBMAP population was obtained by crossing three Guadyerbas Iberian boars with 31 Landrace sows (Pérez-Enciso *et al.,* 2000). In the present study, we used 114 animals belonging to the BC1_LD generation of the IBMAP population obtained by crossing five F1 boars with 26 Landrace sows. Animals were fed *ad libitum* with a cereal-based commercial diet and slaughtered at 179.8 ± 2.6 days. Animal care and procedures were performed following national and institutional guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries). Samples of the *Longissimus dorsi* muscle were collected, snap frozen in liquid nitrogen and stored at -80ºC until further RNA isolation.

Genotyping

A total of 197 animals from the BC1_LD (160 backcrossed individuals and their respective parents) were genotyped with the Porcine *SNP60 Beadchip* (Illumina), following the Infinium HD Assay Ultra protocol (Illumina) (Ramos *et al.,* 2009). Raw data was visualized with GenomeStudio software (Illumina) and trimmed for high genotyping quality (call rate > 0.99). Markers with minor allele frequency (MAF) > 5% and animals with missing genotypes < 5% were retained. After the quality control filter, a subset of 40,586 SNPs remained.

Gene expression profiling

Total RNA was isolated from the *Longissimus dorsi* muscle of 114 samples with RiboPure Isolation of High Quality Total RNA (Ambion, Austin, TX, USA). Total RNA was

quantified in a NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE, USA) and Qubit (Invitrogen, Carlsbad, CA, USA). The RNA was converted to cDNA using the *High-Capacity cDNA Reverse Transcription* (*Applied Biosystems*). The cDNA samples were loaded into a *Dynamic Array* 48.48 chip in a BioMark system (Fluidigm; San Francisco, CA, USA) through a NanoFlex controller. For this experiment, the expressed levels of 48 genes were analyzed and a total of 45 target genes were normalized using the two most stable housekeeping genes (*ACTB* and *TBP*). Primers used for the analyses are detailed in Table S6. Data was collected using the Fluidigm Real-Time PCR analysis software 3.0.2 (Fluidigm) and analyzed using the DAG expression software 1.0.4.11 (Ballester *et al.,* 2013) applying the relative standard curve method (see Applied Biosystems user bulletin #2). Analyses were performed using the normalized gene expression levels of each sample and assay. The animals showing abnormal gene expression levels (outliers) were removed and data obtained was normalized if necessary performing $\log_2$ transformation of the NQ value. We also tested the sex effect by using a linear model with R program (Ihaka & Gentleman, 1996).

Gene expression association analysis

An eGWAS was also performed using the genotypes of BC1_LD animals and the expression values from muscle. A mixed model was employed implemented on Qxpak 5.0 (Pérez-Enciso & Misztal, 2011):

$$y_{ijkl} = Sex_i + Batch_j + \beta c_k + \lambda_k a_l + u_k + e_{ijkl},$$

in which y_{ijkl} was the k^{th} individual record, sex (two levels) and batch (five levels) were fixed effects, ß is a covariate coefficient with *c* being carcass weight, λ_k was a -1, 0, +1 indicator variable depending on the k^{th} individual's genotype for the l^{th} SNP, a_l represents the additive effect associated with the l^{th} SNP, u_k is the infinitesimal genetic effect with random distribution $N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ where **A** is the numerator of the pedigree-based relationship matrix and e_{ijkl} the residual.

To correct for multiple testing, the false discovery rate (FDR) was calculated with the q-value library of R package setting the threshold at q-value ≤ 0.05 (Ihaka & Gentleman, 1996; Storey & Tibshirani, 2003).

The identified eSNPs were classified as *cis* when they were located within 1Mb from the analyzed gene and as *trans* when they were located elsewhere of the genome. The number of significant eSNPs belonging to the same interval was considered among associated eSNPs less than 10Mb apart.

Gene annotation and functional analysis

The significantly associated eSNPs were mapped in the Sscrofa10.2 assembly and were annotated with the Ensembl Genes 78 Database using VEP software (McLaren *et al.,* 2010). The genomic eQTL intervals considering ±1Mb around the candidate chromosomal regions were annotated using BioMart software [http://www.biomart.org].

The Core Analysis function included in the Ingenuity Pathway Analysis (IPA; ingenuity Systems) software [http://www.ingenuity.com] and the Genomatix Pathway System (GePS) (Release 2.8.0) in the Genomatix software suite [https://www.genomatix.de/] were used to perform the functional analysis of genes mapped in the 18 eQTLs regions. Specifically, the IPA software was used for data interpretation in the context of biological processes, pathways, networks, and upstream regulators. All information generated in this software is derived from the Ingenuity Pathway Knowledge Base (IPKB), which is based on functions and interactions of genes published in the literature. Genomatix was used to retrieve additional information of gene functions, interactions and upstream regulators based on literature. Furthermore, information from Mouse Genome Informatics Database [MGI; http://www.informatics.jax.org] and Genecards [http://www.genecards.org] was used to identify gene functions affecting the analyzed phenotypes. For the lipid-related genes, the Gene Expression Atlas [http://www.ebi.ac.uk/gxa] was used to determine whether they were expressed in muscle or not.

Finally, an *in-silico* identification of transcription factor binding sites in the promoter region of the 256 annotated genes was performed. For this analysis we used iRegulon (Janky *et al.,* 2014), which relies on the analysis of the motif enrichment for a transcription factor in the gene set using databases of nearly 10,000 TF motifs and 1,000 ChIP-seq data sets or ''tracks''.

Correlation of gene expression and phenotypes

Correlations were performed among gene expression of the 45 genes to explore the relationship between genes. Furthermore, pairwise correlations among gene expression and FA composition percentages in muscle (Ramayo-Caldas *et al.,* 2012) were carried out to explore the relationships between gene expression and phenotypes. All values were normalized applying the $\log_2$ of raw data if necessary. Afterwards, gene expression was corrected by sex (two levels) and batch (five levels) effects, whereas FA composition percentages were corrected by sex, batch and carcass weight. The remaining residuals of the phenotypes and gene expression values corrected for the corresponding effects were used to obtain the pairwise correlations. The hierarchical clustering option of PermutMatrix software was used to visualize the results of both traits and genes (Caraux & Pinloche 2005).

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

JMF and MB conceived and designed the experiment. JMF was the principal investigator of the project. INIA and JMF collected the samples. MB and APO chose genes, MB and AC designed the primers and MB carried out the Real-Time PCR analyses. MB and MR analyzed the data. MR and APO performed the eGWAS. MB carried out the functional analysis. JMF, MB and APO drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors gratefully acknowledge J.L. Noguera (Institut de Recerca i Tecnologia Agroalimentàries; IRTA) for the animal material contribution. This work was funded by the Ministerio de Economía y Competitividad (MINECO AGL2011-29821-C02 and AGL2014-56369-C2). A. Puig-Oliveras was funded by a "Personal Investigador en Formació" (PIF) grant from Universitat Autònoma de Barcelona (458-01-1/2011).

References

Agulleiro MJ, Sánchez E, Leal E, Cortés R, Fernández-Durán B, Guillot R, *et al.* (2013). Molecular characterization and functional regulation of melanocortin 2 receptor (MC2R) in the Sea Bass. A putative role in the adaptation to stress. Plos One 8(5):e65450.

Ballester M, Cordón R, Folch JM (2013). DAG Expression: high-throughput gene expression analysis on real-time PCR data using standard curves for relative quantification. Plos One 8(11):e80385.

Betz MJ, Hatiboglu N, Mauracher B, Hadaschik D, Sauter A, Demmelmair H, *et al.* (2012). Mc2 receptor knockdown modulates differentiation and lipid composition in adipocytes. Horm Metab Res 44(9):670-5.

Calero G, Gupta P, Nonato MC, Tandel S, Biehl ER, Hofmann SL, *et al.* (2003). The cristal structure of palmitoyl protein thioesterase-2 (PPT2) reveals the basis for divergent substrate specificities of the two lysosomal thioesterases, PPT1 and PPT2. The Journal of Biological Chemistry 278:37957-37964.

Caraux G, Pinloche S (2005). PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics 21:1280-1281.

Chagnon YC, Chen WJ, Perusse L, Chagnon M, Nadeau A, Wilkison WO, *et al.* (1997). Linkage and association studies between the melanocortin receptors 4 and 5 genes and obesity-related phenotypes in the Quebec Family Study. Molecular Medicine 3:663-73.

Clop A, Ovilo C, Pérez-Enciso M, Cercos A, Tomas A, Fernández A, *et al.* (2003) Detection of QTL affecting fatty acid composition in pig. Mamm Genome 14(9):650-6.

Corominas J, Ramayo-Caldas Y, Castelló A, Muñoz M, Ibáñez-Escriche N, Folch JM, *et al.* (2012). Evaluation of the porcine ACSL4 gene as a candidate gene for meat quality traits in pigs. Anim Genet 43(6):714-20.

Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Pérez-Montarelo D, Noguera JL, Folch JM, *et al.* (2013). Polymorphism in the ELOVL6 gene is associated with major QTL effect on fatty acid composition in pigs. Plos One 8(1):e53687.

Datta K, Jinping L, Bhattacharya R, Gasparian L, Wang E, Mukhopadhyay D (2004). Protein kinase C ζ transactivates hypoxia-inducible factor a by promoting its association with p300 in renal cancer. The Journal of Cancer Research 64 :456.

Dobson MG, Redfern CPF, Unwin N, Weaver JU (2001). The N363S polymorphism of the glucocorticoid receptor: potential contribution to central obesity in men and lack of association with other risk factors of coronary heart disease and diabetes mellitus. The Journal of Clinical Endocrinology & Metabolsim 86(5):2270-2274.

Erlmand P, Schmid S, Horenkamp FA, Geyer M, Pomorski TG, Olayioye MA (2009). DLC1 activation requires interaction through a polybasic region preceding the RhoGAP domain. Mol Biol Cell 15:20(20):4400-4411

Estellé J, Mercadé A, Noguera JL, Pérez-Enciso M, Óvilo C, Sánchez A, *et al.* (2005). Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross. J Anim Sci 83(12):2723-8.

Estellé J, Pérez-Enciso M, Mercadé A, Varona L, Alves E, Sánchez A, *et al.* (2006). Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross. Anim Genet 37:589–591.

Estellé J, Gil F, Vázquez JM, Latorre R, Ramírez G, Barragán MC, *et al.* (2008). A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. J Anim Sci 86:3290-3299.

Estellé J, Fernández AI, Pérez-Enciso M, Fernández A, Rodríguez C, Sánchez A, *et al.* (2009). A non-synonymous mutation in a conserved site of the MTTP gene is strongly associated with protein activity and fatty acid profile in pigs. Animal Genetics 40:813-20.

Estévez M, Morcuende D, Cava R (2003). Physico-chemical characteristics of M. Longissimus dorsi from three lines of free-range reared Iberian pigs slaughtered at 90 kg live-weight and commercial pigs: a comparative study. Meat Sci 64(4):499-506.

Fernández AI, Pérez-Montarelo D, Barragán C, Ramayo-Caldas Y, Ibáñez-Escriche N, Castelló A, *et al.* (2012). Genome-wide linkage analysis of QTL for growth and body composition employing the Porcine SNP60 BeadChip. BMC genetics 13:41.

Fontanesi L, Speroni C, Buttazzoni L, Scotti E, Dall'Olio S, Costa NL, *et al.* (2010). The insulin-like growth factor 2 (IGF2) gene intron3-g.3072G>A polymorphism is not the only Sus scrofa chromosome 2p mutation affecting meat production and carcass traits in pigs: evidence from the effects of a cathepsin D (CTSD) gene polymorphism. J Anim Sci 88(7):2235-45.

Fu MH, Maher AC, Hamadeh JM, Ye C, Tarnopolsky MA (2009). Exercise, sex, menstrual cycle phase, and 17β-estradiol influence metabolism-related genes in human skeletal muscle. Physiol Genomics 40:34-47.

Gilad Y, Rifkin SA, Pritchard JK (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24(8):408-415.

Harmon GS, Dumlao DS, Ng DT, Barret KE, Dennis EA, Dong H, *et al*. (2010). Pharmacological correction of a defect in PPAR-γ signaling ameliorates disease severity in Cftr-deficient mice. Nature medicine 16(3):313-318.

Herz J, Chen Y, Masiulis I, Zhou L (2009). Expanding functions of lipoprotein receptors. J Lipid Res 50:S287-S292.

Hong J, Kim D, Cho K, Sa S, Choi S, Kim Y, *et al.* (2015). Effects of genetic variants for the swine FABP3, HMGA1, MC4R, IGF2, and FABP4 genes on fatty acid composition. Meat Sci 110:46-51.

Hung CW, Aoh QL, Joglekar AP, Payne GS, Duncan MC (2012). Adaptor autoregulation promotes coordinated binding within clathrin coats. J Biol Chem 287(21):17398-407.

Hutley L, Shurety W, Newell F, McGeary R, Pelton N, Grant J, *et al.* (2004). Fibroblast Growth Factor 1 a key regulator of human adipogenesis. Diabetes 53(12):3097-106.

Iatan I, Choi HY, Ruel I, Reddy MVPL, Kil H, Lee J, *et al.* (2014). The WWOX gene modulates high-density lipoprotein and lipid metabolism. Circ Cardiovasc Genet 7:491-504.

Ihaka R, Gentleman R (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5(3):299-314.

Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, *et al.* (2014). iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. PLoS Comput Biol 10(7): e1003731.

Jin SH, Yang JH, Shin BY, Seo K, Shin SM, Cho IJ, *et al.* (2013). Resveratrol inhibits LXRα-dependent hepatic lipogenesis through novel antioxidant Sestrin2 gene induction. Toxicol Appl Pharmacol 271(1):95-105.

Jing E, Emanuelli B, Hirschey MD, Boucher J, Lee KY, Lombard D, *et al.* (2011). Sirtuin-3 (SIRT3) regulates skeletal muscle metabolism and insulin signaling via altered mitochondrial oxidation and reactive oxygen species production. PNAS 108(35):14608-14613.

Kapushesky M, Emam I, Holloway E, Kurnosov P, Zorin A, Malone J, *et al.* (2010). Gene expression atlas at the European bioinformatics institute. Nucleic Acids Res 38:D690-D698.

Kobayashi N, Ueki K, Okazaki Y, Iwane A, Kubota N, Ohsugi M, *et al.* (2001). Blockade of class IB phosphoinositide-3 kinase ameliorates obesity-induced inflammation and insulin resistance. PNAS 108(14):5753-5758.

Liu B, Agras K, Willingham E, Vilela MLB, Baskin LS (2006). Activating transcription factor 3 is estrogen-responsive in utero and upregulated during sexual differentiation. Hormone Research in Pediatrics 65(5):217-222.

Liu D, Sartor MA, Nader GA, Gutmann L, Treutelaar MK, Pistilli EE, *et al.* (2010). Skeletal muscle gene expression in response to resistance exercise: sex specific regulation. BMC Genomics 11:659.

McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26(16):2069-70.

Mercadé A, Estellé J, Noguera JL, Folch JM, Varona L, Silió L, *et al*. (2005a). On growth fatness and form: A further look at porcine Chromosome 4 in an Iberian × Landrace cross. Mamm Genome 16:374–382.

Mercadé A, Sánchez A, Folch JM (2005b). Exclusion of the acyl CoA:diacylglycerol acyltransferase 1 gene (DGAT1) as a candidate for a fatty acid composition QTL on porcine chromosome 4. J Animal Breeding and Genetics 122(3):161-4.

Mercadé A, Pérez-Enciso M, Varona L, Alves E, Noguera JL, Sánchez A, *et al.* (2006). Adipocyte fatty-acid binding protein is closely associated to the porcine FAT1 locus on chromosome 4. J Anim Sci 84:2907–2913.

Michas G, Micha R, Zampelas A (2014). Dietary fats and cardiovascular disease: putting together the pieces of a complicated puzzle. Atherosclerosis 234:320-328.

Mirza R, Hayasaka S, Takagishi Y, Kambe F, Ohmori S, Maki K, *et al.* (2006). DHCR24 gene knockout mice demonstrate lethal dermopathy with differentiation and maturation defects in the epidermis. J Invest Dermatol 126(3):638-47.

Montooth K, Siebenthall KT, Clark A (2006). Membrane lipid physiology and toxin catabolism underlie ethanol and acetic acid tolerance in Drosophila melanogaster. The Journal of Experimental Biology 209, 3837-3850.

Mormeneo E, Jimenez-Mallebrera C, Palomer X, De Nigris V, Vázquez-Carrera M, Orozco A, *et al.* (2012). PGC-1α induces mitochondrial and myokine transcriptional programs and lipid droplet and glycogen accumulation in cultured human skeletal muscle cells. Plos One 7(1):e29985.

Muñoz M, Alves E, Corominas J, Folch JM, Casellas J, Noguera JL, *et al.* (2011). Survey of SSC12 regions affecting fatty acid composition of intramuscular fat using high-density SNP data. Front Genet 2:101.

Muñoz M, Rodríguez MC, Alves E, Folch JM, Ibañez-Escriche N, Silió L, *et al.* (2013). Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. BMC Genomics 14:845.

Newsom SA, Boyle KE, Friedman J (2013). Sirtuin 3: A major control point for obesity-related metabolic diseases?. Drug Discov Today Dis Mech 10(1-2):e35-e40.

Óvilo C, Pérez-Enciso M, Barragán C, Clop A, Rodríguez C, Oliver MA, *et al.* (2000). A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6. Mammalian Genome 11:344-6.

Óvilo C, Oliver A, Noguera JL, Clop A, Barragán C, Varona L, Rodríguez C, *et al.* (2002). Test for positional candidate genes for body composition on pig chromosome 6. Genet Sel Evol 34:465-479.

O'Neill HM, Holloway GP, Steinberg GR (2013). AMPK regulation of fatty acid metabolism and mitochondrial biogenesis: Implications for obesity. Molecular and Cellular Endocrinology 366:135-151.

Parl A, Mitchell SL, Clay HB, Reiss S, Li Z, Murdock DG (2013).The mitochondrial fatty acid synthesis (mtFASII) pathway is capable of mediating nuclear-mitochondrial cross talk through the PPAR system of transcriptional activation. Biochemical and Biophysical Research Communications 441(2):418-424.

Pérez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, Folch JM, *et al.* (2000). A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross. J Anim Sci 78:2525–2531.

Pérez-Enciso M, Misztal I (2011). Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 12:202.

Petzold KM, Naumann H, Spagnoli FM (2013). Rho signalling restriction by the Rho GAP Stard13 integrates growth and morphogenesis in the pancreas. Development 140:126-135.

Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, *et al.* (2014a). Differences in muscle transcriptome among pigs phenotipically extreme for fatty acid composition. Plos One 9(7):e103668.

Puig-Oliveras A, Ballester M, Corominas J, Revilla M, Estellé J, Fernández AI, *et al.* (2014b). A co-association network analysis of the genetic determination of pig conformation growth and fatness. Plos One 9(12):e114862.

Raddatz K, Turner N, Frangioudakis G, Liao BM, Pedersen DJ, Cantley J, *et al.* (2011). Time-dependent effects of Prkce deletion on glucose homeostasis and hepatic lipid metabolism on dietary lipid oversupply in mice. Diabetologia 54:1447-1456.

Ramayo-Caldas Y, Mercadé A, Castelló A, Yang B, Rodríguez C, Alves E, *et al.* (2012). Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. J Anim Sci 90:2883–2893.

Ramayo-Caldas Y, Ballester M, Fortes M, Esteve-Codina A, Castello A, *et al.* (2014) From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. BMC Genomics 15:232.

Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, *et al.* (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.

Rao DC (2008). An Overview of the Genetic Dissection of Complex Traits. Advances in Genetics 60:3–34.

Remark AS (1995). R94: A remark on algorithm AS 181: The W-test for normality. Journal of the Royal Statistical Society. Series C (Applied Statistics), 44(4):457-551.

Renvoisé B, Stadler J, Singh R, Bakowska JC, Blackstone C (2012). Spg20-/- mice reveal multimodal functions for Troyer syndrome protein spartin in lipid droplet maintenance, cytokinesis and BMP signaling. Hum Mol Genet 21(16):3604:18.

Revilla M, Ramayo-Caldas Y, Castelló A, Corominas J, Puig-Oliveras A, Ibáñez-Escriche N, *et al.* (2014). New insight into the SSC8 genetic determination of fatty acid composition in pigs. Genetics Selection Evolution 46:28.

Rubin CJ, Megens HJ, Barrio AM, Maqbool K, Sayyab S, Schwochow D, *et al.* (2012). Strong signatures of selection in the domestic pig genome. PNAS 109(48):19529-19536.

Rune A, Salehzadeh F, Szekeres F, Kühn I, Osler ME, Al-Khalili L (2009). Evidence against a sexual dimorphism in glucose and fatty acid metabolism in skeletal muscle cultures from age-atched men and post-menopausal women. Acta Physiol 197:207-215.

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, *et al.* (2003). Genetics of gene expression surveyed in maize, human and man. Nature 422(6929):297-302.

Schug TT, Li X (2011). Sirtuin 1 in lipid metabolism and obesity. Ann Med 43(3):198-211.

Schwab CR, Baas TJ, Stalder KJ, Mabry JW (2006). Effect on long-term selection for leanness on meat and eating quality traits in Duroc swine. J Anim Sci 84:1577-1583.

Semba S, Itoh N, Ito M, Youssef EM, Harada M, Moriya T, *et al.* (2002). Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-OH kinase, by CpG hypermethylation in human colorectal carcinoma. Clinical Cancer Research 8:2824-3831.

Serra X, Gil F, Pérez-Enciso M, Oliver M, Vázquez J, Gispert M, *et al.* (1998). A comparison of carcass meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Livest Prod Sci 56:215–223.

Shi HB, Luo J, Yao DW, Zhu JJ, Xu HF, Shi HP, *et al.* (2013). Peroxisome proliferator-activated receptor-γ stimulates the synthesis of monounsaturated fatty acids in dairy goat mammary epithelial cells via the control of stearoyl-coenzyme A desaturase. J Dairy Sci 96:7844-7853.

Smith RD, Lupashin VV (2008). Role of the conserved oligomeric Golgi (COG) complex in protein glycosylation. Carbohydrate Research 343(12):2024-31.

Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. PNAS 100(16):9440-9445.

Tao R, Gong J, Luo X, Zang M, Guo W, Wen R, *et al.* (2010). AMPK exerts dual regulatory effects on the PI3K pathway. Journal of Molecular Signaling 5:1.

Thorsell AG, Lee WH, Persson C, Siponen MI, Nilsson M, *et al.* (2011). Comparative structural analysis of lipid binding START domains. Plos One 6(6):e19521.

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collete C, Moreau L, *et al*. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425, 832-836.

Virtue S, Masoodi M, de Weijer BA, van Eijk M, Mok CY, Eiden M, *et al*. (2015). Prostaglandin profiling reveals a role for haematopoietic prostaglandin D synthase in adipose tissue macrophage polarisation in mice and humans. Int J Obes 39(7):1151-60.

Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, Richardson RI, *et al*. (2008). Fat deposition fatty acid composition and meat quality: A review. Meat Sci 78:343-358.

Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, *et al.* (2011). Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. Plos One 6(8):e23406.

Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, *et al*. (2012). AnimalTFDB: a comprehensive animal transcription factor database. Nucl Acids Res 40:D144-D149.

Zhen Y, Sørensen V, Skjerpen CS, Haugsten EM, Jin Y, Wälchli S, *et al.* (2012). Nuclear import of exogenous FGF1 requires the ER-Protein LRRC59 and the importins Kpnα1 and Kpnβ1. Traffic 13:650-664.

Zhou L, Xu L, Ye J, Li D, Wang W, Li X, *et al.* (2012). Cidea promotes hepatic steatosis by sensing dietary fatty acids. Hepatology 65(1):96-107.

Figure Legends

Figure 1. Hierarchical cluster of correlations among gene expression levels (RQ) of the 45 genes and fatty acid content in muscle.

Rows : - Objective function : R=0.500
- Sum of all pairwise distances of neighboring rows (path length): S=14.291
- Linkage rule: McQuitty's criteria

Columns : - Objective function : R=0.538
- Sum of all pairwise distances of neighboring columns (path length): S=10.458
- Linkage rule: McQuitty's criteria

Dissimilarity : - Euclidean distance

The colors scale:

Min = -0.30 0.00 Max = 0.30

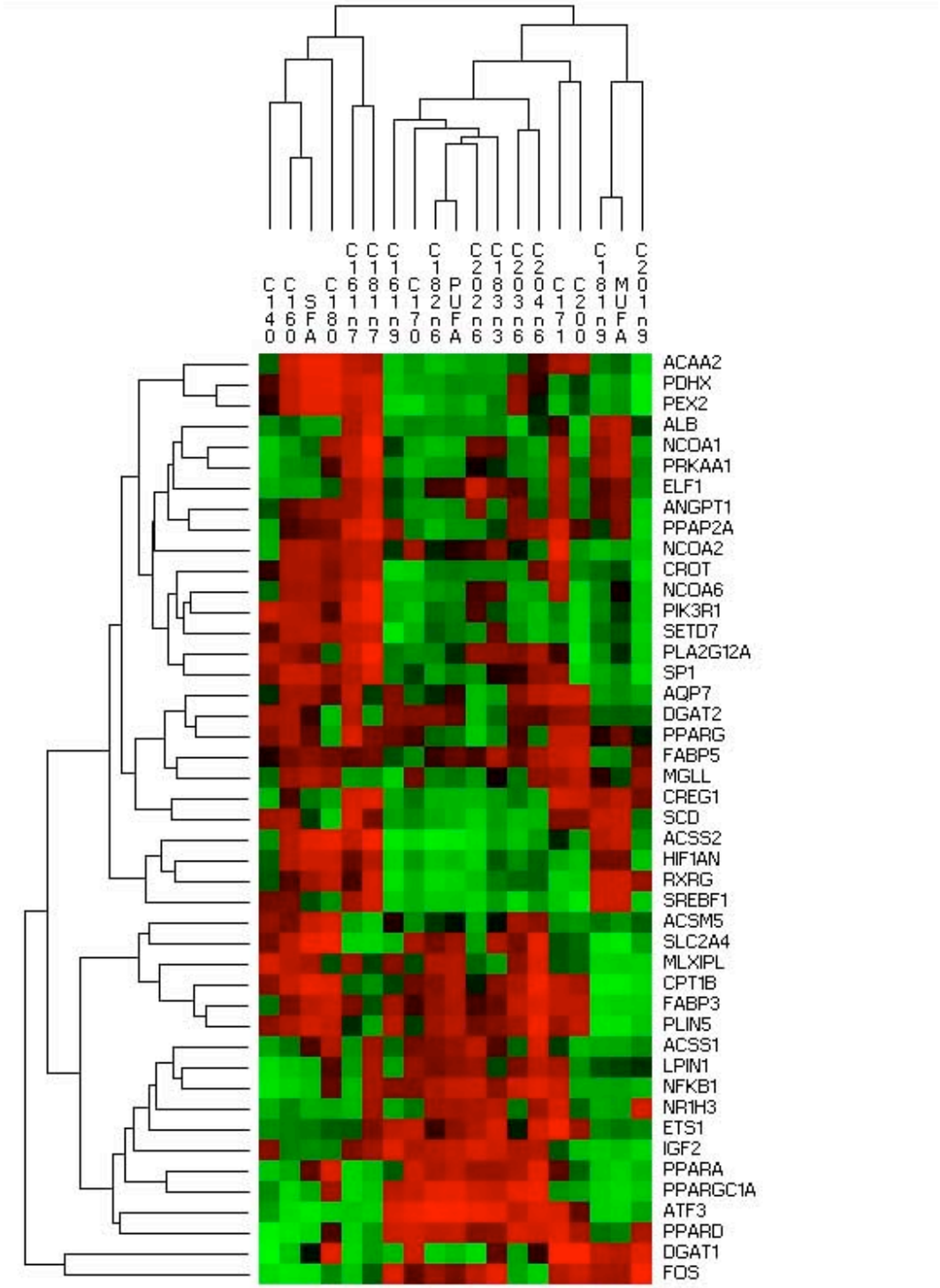
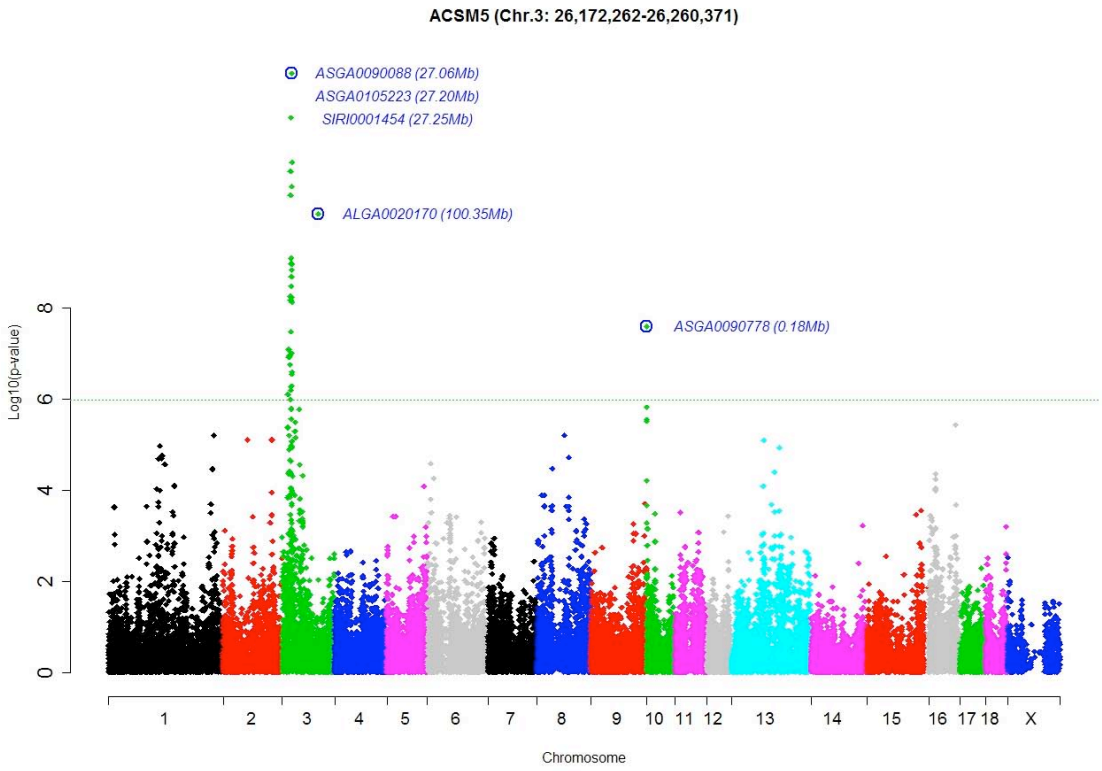
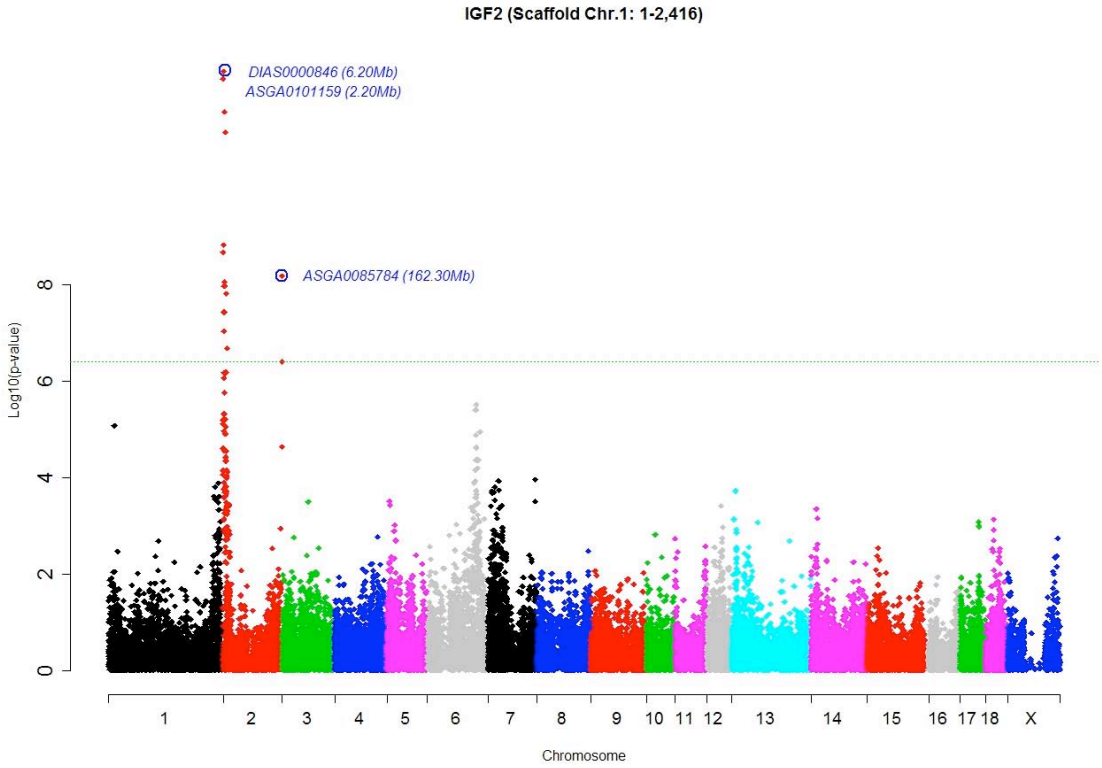


Figure 2. GWAS plot of *ACSM5, IGF2* and *MGLL* gene expression in muscle tissue. Positions in Mb are relative to *Sscrofa10.2* assembly of the pig genome. Horizontal dashed lines indicate the chromosome wide significance level (FDR-based q-value ≤ 0.001). Plot of eGWAS for (A) *ACSM5* gene expression in muscle (B) *IGF2* gene expression in muscle (C) *MGLL* gene expression in muscle.

A)



B)


IGF2 (Scaffold Chr.1: 1-2,416)
DIAS0000846 (6.20Mb)
ASGA0101159 (2.20Mb)
ASGA0085784 (162.30Mb)
Log10(p-value)
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

C)

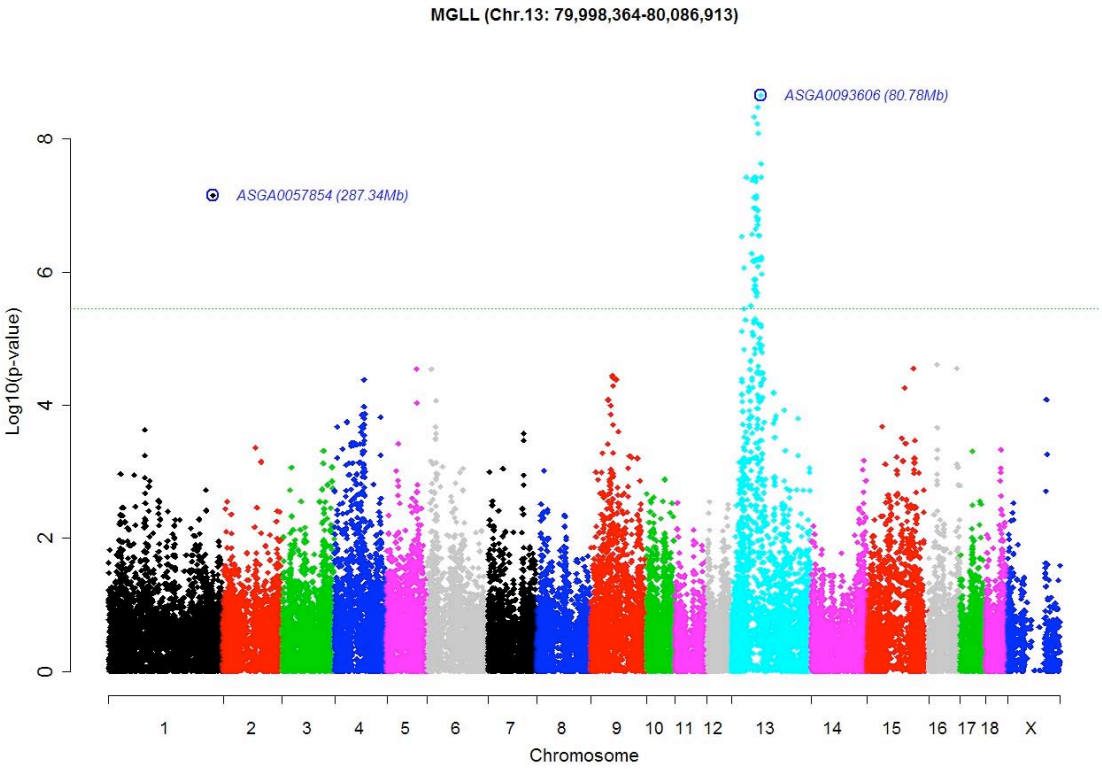
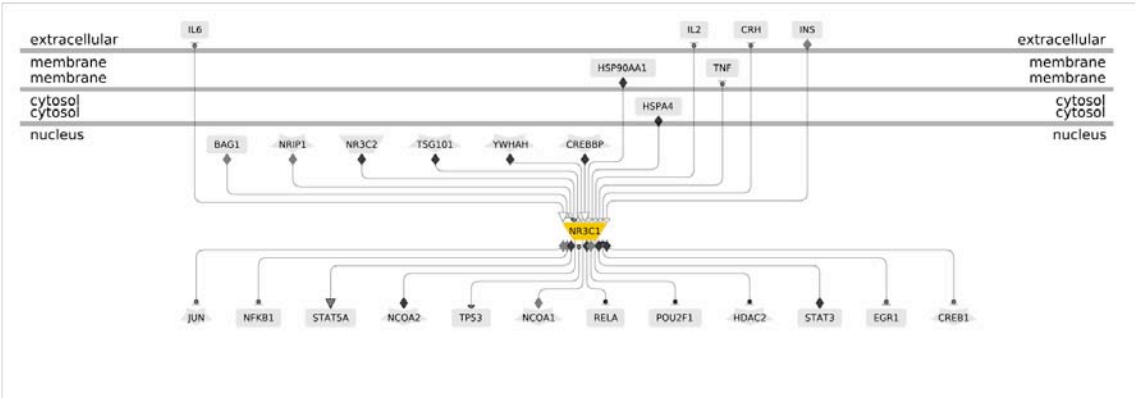

MGLL (Chr.13: 79,998,364-80,086,913)
ASGA0093606 (80.78Mb)
ASGA0057854 (287.34Mb)
Log10(p-value)
0
2
4
6
8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
X
Chromosome

Figure 3. Literature-based network of genes interacting with NR3C1 transcription factor generated by Genomatix.



Tables

Table 1. Significant eQTLs identified.

No.	Gene	most significant SNPs in eQTL	p-value	q-value	eQTL interval[1]	No. SNPs[2]	Type of eQTL
1	*ACSM5*	ASGA0090088, ASGA0105223, SIRI0001454	7.12E-14	9.29E-10	3:16487448-27941321	43	*CIS*
2	*ACSM5*	ALGA0020170	8.58E-11	3.73E-07	3:100347076	1	*TRANS*
3	*ACSM5*	ASGA0090778	2.55E-08	3.44E-05	10:175359	1	*TRANS*
4	*CROT*	ALGA0046590	4.01E-07	9.77E-03	8:15870120-15912410	3	*TRANS*
5	*FABP3*	ALGA0063896	1.26E-06	4.47E-02	11:81364018-82160789	2	*TRANS*
6	*FOS*	H3GA0031293	1.22E-09	4.95E-05	11:6736568-11968518	6	*TRANS*
7	*HIF1AN, PLA2G12A*	H3GA0028012, ASGA0044215, ALGA0117195	2.91E-07 4.48E-06	3.92E-03 2.74E-02	9:117742788-117851340	3	*TRANS*
8	*IGF2*	ASGA0101159, DIAS0000846	<1.0E-25	<1.0E-25	2:16416-11175095	17	*CIS**
9	*IGF2*	ASGA0085784	6.63E-09	2.45E-05	2:162088043-162298086	2	*TRANS*
10	*MGLL*	ASGA0057854	7.06E-08	6.28E-05	1:287348708	1	*TRANS*
11	*MGLL*	ASGA0093606	2.20E-09	2.61E-05	13:27954256-82589660	147	*CIS*
12	*NCOA1*	ALGA0016576, MARC0045025, MARC0087200	2.60E-06	3.49E-02	2:146195530-146722998	3	*TRANS*
13	*PIK3R1*	ALGA0117336	1.24E-06	3.02E-02	6:143952619-145605378	3	*TRANS*
14	*PLA2G12A*	ALGA0113789	1.96E-08	7.84E-04	6:10020702	1	*TRANS*
15	*PLA2G12A*	ALGA0103867	3.41E-06	2.74E-02	6:79924228	1	*TRANS*
16	*PLA2G12A*	ASGA0039774	1.34E-06	2.68E-02	8:128899782-136113975	3	*TRANS*
17	*PPARA*	M1GA0003328	3.25E-06	3.30E-02	2:150634202	1	*TRANS*
18	*PPARA*	MARC0074986, DIAS0004325, CASI0006620	6.17E-07	6.17E-07	6:89986075-90352248	3	*TRANS*

[1] Chromosomal location is given according to the Sscrofa10.2 assembly coordinates. Positions are relative to the significant eQTL interval start and end. Lengths are given in base pairs.

[2] Number of significant eSNPs within the eQTL interval.

* Approximate location of the gene described by Fontanesi *et al.* (2010).

Supporting Information

Figure S1. Comparison between males and females of muscle gene expression levels of 45 lipid-related genes. Data represents means ± SEM. Significant differences between sexes are indicated as * P ≤ 0.05, ** P ≤ 0.01, ** P ≤ 0.001.

Figure S2. Network of the lipid metabolic process function obtained by Genomatix.

Figure S3. Network (score 44) generated by IPA of 25 focus genes corresponding to the energy production, small molecule biochemistry and drug metabolism functions.

Table S1. Descriptive statistics including mean and standard deviation (SD) of intramuscular fat (IMF), fatty acid (FA) composition and fatty acid indices of the BC1_LD animals analyzed.

Table S2. Description of the 241 eSNPs identified as significantly associated with gene expression.

Table S3. Gene annotation within the eQTL intervals. Annotation was performed by considering for *trans*-eQTLs the eQTL interval ±1Mb; whereas for *cis*-eQTLs only the studied gene was selected (*ACSM5, IGF2,* and *MGLL*).

Table S4. Top functional networks identified with IPA based on the list of annotated genes mapping within the 18 eQTLs.

Table S5. Positional concordance among the 241 eSNPs associated with gene expression and the QTLs described in the pig QTL database for fatness and fat composition related traits.

Table S6. Primers used for the analyses of gene expression of the 48 genes by Real-Time PCR.