



Universitat Autònoma de Barcelona

Facultat de Veterinària

Departament de Ciència Animal i dels Aliments

**Dissecting the Genetic Basis of Pig Intramuscular Fatty
Acid Composition**

Ph. D, Thesis

YULIAXIS RAMAYO CALDAS

Supervisors

Dr. Josep M. Folch Albareda

Dr. Miguel Pérez Enciso

BELLATERRA, JULY 2013

“I know one thing: that I know nothing”

- Socrates

“Solo se ve bien con el corazón.

Lo esencial es invisible a los ojos”

- Antoine de Saint-Exupéry

SUMMARY

Pigs (*Sus scrofa*) are relevant to humans, both as a source of food and as an animal model for scientific progress. Technological, nutritional and organoleptic properties of pork meat quality are highly dependent on lipid content and fatty acid (FA) composition. The molecular processes controlling FA composition and metabolism are highly interconnected and not fully understood. Elucidating these molecular processes will aid the technological development towards the improvement of pork meat quality and increase knowledge of FA metabolism underpinning metabolic diseases in humans. This thesis deals with FA metabolism in pigs, analyzed from different perspectives. We have described for the first time a map of Copy Number Variant Regions in the swine genome. It is based on whole genome SNP genotyping chips and some of them may have an effect on FA metabolism.

A Genome-Wide Association Study (GWAS) approach was carried out in an Iberian x Landrace backcross across 32 traits related to intramuscular FA composition and 813 associated SNPs, distributed in 43 chromosomal regions, were identified. Besides, 53.5% of these chromosomal intervals had been reported to contain QTL for lipid traits in previous studies, being particular consistent the regions detected on chromosomes SSC4, SSC8 and SSC16. The liver transcriptome of two groups of phenotypically extreme pigs for intramuscular FA composition were sequenced using RNA-Seq. A differential expression analysis identified 55 protein-coding genes differentially-expressed (DE) between groups. Functional analysis revealed that these genes belong to biological functions, canonical pathways and three gene networks related to lipid and FA metabolism. Moreover, a connection was observed among the top three networks, suggesting that the DE genes detected in this study are linked and play an important and coordinate role in lipid and FA metabolism.

The complexity and high interconnectivity of the molecular processes controlling FA metabolism identified by our GWAS and RNA-Seq results, motivated us to use a holistic gene network approach based on SNP-to-SNP co-association analysis. Supporting evidence for co-association network predictions were confirmed at tissue-specific manner by gene co-expression analysis in adipose and liver tissues. To the best of our knowledge, we reported for the first time in a vertebrate, the observation that interacting loci could jointly regulate the co-expression patterns of pairs of genes.

The analysis of the topological properties of both the co-association and co-expression predicted gene networks, allowed the identification of key transcription factors (TF), candidate genes and metabolic pathways that are likely being determining meat quality and FA composition, as well as controlling energy homeostasis in pigs. Future studies targeting these genes, their pathways and interactions will continue to expand our knowledge of molecular control of FA metabolism and it might lead to the discovery of functional relevant mutations, unfolding new strategies for improve pork meat quality.

Keywords: pig, meat quality, intramuscular fatty acid composition, lipid metabolism, candidate genes, copy number variant regions, GWAS, RNA-Seq, gene network, transcription factors, SNPs, co-association, co-expression.

RESUMEN

El cerdo (*Sus scrofa domestica*) constituye una de las principales fuentes de carne para la humanidad, y es también un excelente modelo animal para el estudio de enfermedades metabólicas en humanos. Los ácidos grasos (AG) juegan un papel importante actuando como moléculas de señalización celular en diversas rutas metabólicas asociadas a este tipo de patologías y además en la determinación de la calidad organoléptica y nutricional de la carne porcina. Como consecuencia, descifrar la base molecular del metabolismo de los AG podría contribuir al conocimiento de la base genética de enfermedades metabólicas humanas, además de favorecer el desarrollo tecnológico para mejorar la calidad de la carne porcina.

El objetivo principal de esta tesis ha sido estudiar la base genética del contenido intramuscular de AG en cerdos. Para ello hemos utilizado diferentes, pero complementarias, aproximaciones analíticas. Inicialmente utilizamos los datos genotípicos del chip porcino de 62.612 SNPs (Illumina) e identificamos 49 regiones con variación en el número de copias (CNVRs). Posteriormente, estudios de asociación masivos (GWAS) entre 32 fenotipos relacionados con el perfil intramuscular de AG y los SNPs del chip de Illumina nos permitieron identificar 813 SNPs que distribuidos en 43 regiones están significativamente asociados a 15 de los fenotipos analizados. Cabe destacar que el 53.5% de las regiones identificadas contienen QTLs asociados a caracteres lipídicos. Esta observación sugiere que algunos de estos QTLs podrían tener un efecto pleiotrópico, especialmente en las regiones identificadas en los cromosomas porcinos SSC4, SCC8 y SSC16.

Al comparar el transcriptoma del hígado entre dos grupos de cerdos con fenotipos extremos para la composición intramuscular de AG se han identificado un total de 55 genes diferencialmente expresados. Análisis funcionales revelan que muchos de estos genes pertenecen a funciones biológicas, rutas metabólicas y redes génicas interconectadas relacionadas con el metabolismo de los lípidos y de los AG.

Los resultados obtenidos en el estudio de asociación y en el análisis del transcriptoma hepático confirman la complejidad e interconexión de los procesos moleculares relacionados con el metabolismo de los AG. Por lo tanto, nos propusimos utilizar un enfoque holístico basado en la co-asociación entre genes y/o SNPs a lo largo de varios fenotipos. Este análisis nos permitió identificar redes génicas, rutas metabólicas y factores de transcripción (TF) relevantes para el metabolismo de los AG. Por primera vez, interacciones génicas predichas en base a la co-asociación entre genes fueron validadas mediante estudios de co-expresión génica en dos tejidos relevantes para el metabolismo de los AG en cerdos.

Los resultados obtenidos muestran la utilidad de la biología de sistemas en el estudio de caracteres complejos. Sin embargo, serán necesarios estudios que integren diferentes disciplinas para establecer estrategias más eficientes que permitan mejorar la calidad de la carne de cerdo.

Palabras claves: cerdos, calidad de la carne, ácidos grasos, metabolismo lipídico, gen candidato, CNV, GWAS, RNA-Seq, red génica, factor de transcripción, SNPs, co-asociación, co-expresión.

CONTENTS

List of Tables.....	9
List of Figures.....	10
List of Publications.....	13
Related publications by the author.....	14
Abbreviations.....	16
Chapter I. General Introduction.....	18
Pork meat quality.....	20
Pig genomics.....	23
Mapping complex traits in pigs.....	27
Chapter II. Objectives.....	36
Chapter III. General Discussion.....	37
Detecting CNVR in the pig Genome.....	38
From GWAS to Gene-Networks through expression and co-expression analysis.....	44
Future directions.....	50
Chapter IV. Conclusions.....	54
Chapter V. References.....	55

Chapter VI. Papers included in the thesis..... 84

Paper I. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip.

Paper II. Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross.

Paper III. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition.

Chapter VII. Annexes. 124

From SNP co-association to gene co-expression: Evidence of pleiotropic expression-QTL in lipid metabolism in pigs.....125

LIST OF TABLES

General Introduction

Table 1. Traits related with the sensorial, technological and nutritional quality of pork meat quality.....	21
Table 2. Assembly and annotation statistics corresponding to the Sscrofa10.2 genome assembly.....	25
Table 3. Functional polymorphisms affecting meat and/or carcass traits in the pig.....	30

General Discussion

Table 4. Summary of independent confirmations of 26 CNVRs	40
Table 5. Positional concordance among candidate genes for FA metabolism located within CNVRs and QTL deposited in the pig QTL database for fatness related traits.....	42

Annexes

Table 6. Description of the ten most connected nodes in the co-association network.....	129
Table 7. Concordance validation between the co-association and the co-expression networks for the best TF trio and in the two tissues.....	137

LIST OF FIGURES

General Introduction

Figure 1. Distribution of pork production among the autonomous communities in Spain..... 19

Figure 2. Distribution of reported QTLs for meat quality traits across the pig genome..... 29

General Discussion

Figure 3. Evidence of functional interactions among the three TF (*NCOA2*, *EP300*, *FHL2*) and TF considered as master regulators of lipid and carbohydrate metabolism..... 47

Figure 4. Nuclear receptors influenced by n-3 FA. FXR, farnesoid X receptor; LXR, liver X receptor; NF κ B, nuclear factor k B; PPAR, peroxysome proliferators activated receptor; RXR, retinoid X receptor; SREBP-1c, sterol regulatory element binding protein 1c..... 53

Annexes

Figure 5. Hierarchical cluster analysis of the 15 phenotypes analyzed in this study: Palmitic acid (C16), Stearic acid (C18), Palmitoleic acid (C16:1n7), Oleic acid (C18:1n9), Linoleic acid (C18:2n6), α -Linolenic acid (C18:3n3), Eicosadienoic acid (C20:2n6), Eicosatrienoic acid (C20:3n6), Arachidonic acid (C20:4n6), Saturated FA (SFA), Monounsaturated FA (MUFA), Polyunsaturated FA (PUFA), Unsaturated indices (UI), Elongase activity (C20/C18), Percentage intramuscular fat (IMF)...128

Figure 6. Co-association network based on the AWM approach. (A) Entire network with 1,096 nodes (i.e., genes or SNPs) and 111,198 interactions. The color spectrum ranges from green to red for low and high density, respectively. (B) Subset of the network showing the best trio of transcription factors: *NCOA2*, *EP300* and *FHL2*. Node color corresponds with the functional classification of the in silico predicted target gene as follows: TF (red), lipid metabolism process (blue), carbohydrate metabolisms (green), development process (orange) and finally, white nodes represent genes with others functional classification. Node shape indicate classification as: diamond (TF involved in lipid metabolism), triangle (TF), ellipse (other genes)...130

Figure 7. Results of the liver differential expression analysis comparing the best TF trio in the Iberian and Landrace breeds.....133

Figure 8. Connections from the co-association network that were confirmed by the co-expression network in liver (A) and adipose (B) tissues. Nodes colors relate to the functional classification of genes as follows: TF (red nodes), lipid metabolism (blue nodes), carbohydrate metabolism (green), development process (orange) and white nodes represent genes with other functional classification. The size of the nodes corresponding to the best trio of transcription factors (*NCOA2*, *EP300* and *FHL2*) has been enlarged to facilitate their location.....136

Figure 9. Sensitivity analysis of the 34 lipid-related TF at different P -values (from $P < 1$ to $P < 10^{-4}$) against the distribution of the total number of TF included in the dataset..... 148

LIST OF PUBLICATIONS

This thesis is based on the work contained in the following published papers:

Paper I **Ramayo-Caldas Y**, Castelló A, Pena RN, Alves E, Mercadé A, Souza CA, Fernández AI, Perez-Enciso M, Folch JM (2010). Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics. 2010 Oct 22;11:593. doi: 10.1186/1471-2164-11-593.

Paper II **Ramayo-Caldas Y**, Mercadé A, Castelló A, Yang B, Rodríguez C, Alves E, Díaz I, Ibáñez-Escriche N, Noguera JL, Pérez-Enciso M, Fernández AI, Folch JM. (2012). Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. J Anim Sci. 2012 Sep;90(9):2883-93. doi: 10.2527/jas.2011-4900.

Paper III **Ramayo-Caldas Y**, Mach N, Esteve-Codina A, Corominas J, Castelló A, Ballester M, Estellé J, Ibáñez-Escriche N, Fernández AI, Pérez-Enciso M, Folch JM. (2012). Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics. 2012 Oct 11;13:547. doi: 10.1186/1471-2164-13-547.

and the recently submitted paper:

Paper IV **Ramayo-Caldas Y**, Ballester M, Fortes M, Esteve-Codina A, Castelló A, Noguera JL, Fernandez AI, Pérez-Enciso M, Reverter A, Folch JM (2013). From SNP co-association to gene co-expression: Evidence of pleiotropic expression-QTL in lipid metabolism in pigs.

RELATED PUBLICATIONS BY THE AUTHOR

(Not included in the thesis)

- J. Corominas, **Y. Ramayo-Caldas**, A. Puig-Oliveras, D. Pérez-Montarelo, J.L. Noguera, J.M. Folch, M. Ballester (2013). Polymorphism in the ELOVL6 Gene Is Associated with a Major QTL Effect on Fatty Acid Composition in Pigs. PLoS ONE 01/2013; 8(1):e53687.

- W. Burgos-Paz, C.A. Souza, H.J. Megens, **Y. Ramayo-Caldas**, M. Melo, C. Lemús-Flores, E. Caal, HW. Soto, R. Martínez, L A. Álvarez, L. Aguirre, V. Iñiguez, MA. Revidatti, O R Martínez-López, S Llambi, A Esteve-Codina, M C Rodríguez, R P M A Crooijmans, SR. Paiva, LB. Schook, MAM. Groenen & M. Pérez-Enciso (2012). Porcine colonization of the Americas: a 60k SNP story. Heredity 12/2012. doi:10.1038/hdy.2012.109.

- D. Pérez-Montarelo, N. J Hudson, A. I Fernández, **Y. Ramayo-Caldas**, B. P Dalrymple, A. Reverter. Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. PLoS ONE 01/2012; 7(9):e46159.

- M. Ballester, Anna Castelló, **Y. Ramayo-Caldas**, J. M Folch. A Quantitative Real-Time PCR Method Using an X-Linked Gene for Sex Typing in Pigs. Molecular Biotechnology 07/2012.

- A. I Fernández, D. Pérez-Montarelo, C. Barragán, **Y. Ramayo-Caldas**, N. Ibáñez-Escriche, A. Castelló, J. L. Noguera, L. Silió, J. M. Folch, M C. Rodríguez. Genomewide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. BMC Genetics 05/2012; 13:41.

- J. Corominas, **Y. Ramayo-Caldas**, A. Castelló, M. Muñoz, N. Ibáñez-Escriche, J.M. Folch, M. Ballester. Evaluation of the porcine ACSL4 gene as a candidate gene for meat quality traits in pigs. Animal Genetics 03/2012; 43(6):714-20.

- M. Muñoz, E. Alves, **Y. Ramayo-Caldas**, J. Casellas, C. Rodríguez, J.M. Folch, L. Silió, A.I. Fernández. Recombination rates across porcine autosomes inferred from highdensity linkage maps. Animal Genetics 12/2011; 43(5):620-3.

ABBREVIATIONS


IBMAP	Iberian x Landrace intercross
IMF	Intramuscular Fat
FA	Fatty acid composition
ALA	α -linolenic acid
LA	Linolenic acid
EDA	Eicosadienoic acid
ETE	Eicosatrienoic acid
AA	Arachidonic acid
ACL	Average Chain Length
SFA	Saturated Fatty Acids
MUFA	Monounsaturated Fatty Acids
PUFA	Polyunsaturated Fatty Acids
PI	Peroxidability index
DBI	Double-bond index
UI	Unsaturated index
PCR	Polymerase chain reaction
qPCR	Quantitative real time PCR
CNV	Copy Number Variant
CNVR	Copy Number Variant Region
SNP	Single Nucleotide Polymorphism
<i>CYP450</i>	Cytochrome P450 gene family
<i>SLC16A7</i>	Solute carrier family 16 member 7
<i>MCT2</i>	Monocarboxylic acid transporter 2
<i>SOX14</i>	SRY (sex determining region Y)-box 14
<i>INSC</i>	Inscuteable homolog
QTL	Quantitative trait locus
h^2	Heritability
GWAS	Genome-wide association study
P	P-value
FDR	False discovery rate
Mb	Megabase
Kb	Kilobase
LD	Linkage disequilibrium
LDLA	Linkage disequilibrium and Linkage analysis
MAF	Minor allele frequency
IPA	Ingenuity Pathways Analysis
RXR	Retinoid X receptors
PXR	Pregnane X receptor
FXR	Farnesoid X receptor
LXR	Liver X receptor
PPAR- α	Peroxisome proliferator-activated receptors alpha
AWM	Association Weight Matrix
GO	Gene Ontology
TF	Transcription factors
Deg	Average degree
AvD _G	Average distance

<i>r</i>	Correlation
FC	Fold chance
<i>NCOA2</i>	Nuclear receptor coactivator 2
<i>EP300</i>	E1A binding protein p300
<i>FHL2</i>	Four and a half LIM domains 2
<i>SREBP-1</i>	Sterol regulatory element binding transcription factor 1
<i>PPARG</i>	Peroxisome proliferator-activated receptor gamma
<i>HNF4-α</i>	Hepatocyte nuclear factor 4, alpha
<i>AR</i>	Androgen receptors
<i>ER</i>	Estrogen receptor
<i>ARNT</i>	Aryl hydrocarbon receptor nuclear translocator
HIF-1	Hypoxia-inducible factor 1 signaling pathway
T2D	Type II diabetes
<i>BCL9</i>	B-cell CLL/lymphoma 9
<i>WNT4</i>	Wingless-type MMTV integration site family, member 4
HPRD	The Human Protein Reference Database
BIND	Biomolecular Object Network Databank
PCIT	Partial Correlation coefficient with Information Theory

CHAPTER I

GENERAL

INTRODUCTION



“All models are wrong, but some are useful”

- George E. P. Box

CHAPTER 1. GENERAL INTRODUCTION

The pig (*Sus scrofa*) is a *Eutherian* mammal member of the *Cetartiodactyla* order and *Suidae* family. Genetic and archaeological evidences suggest that *Sus scrofa* are originally from South East Asia, and were independently domesticated in western Eurasia and East Asia ca. 10 KYA (Groenen et al., 2012; Larson et al., 2005). Modern pigs are common worldwide reflecting their relevance to humans as a source of food. Pork productions constitute an important component of agriculture accounting for more than 40% of the meat produced worldwide (Rothschild et al., 2011). Pig grows rapidly, reproduces at a high rate and is accommodating in its eating habits. These characteristics facilitated the extension of pork production worldwide.

China is so far the major producer, accounting for the 49% of the produced meat, followed by United States of America, Germany and Spain (FAO 2011: <http://faostat.fao.org/>). In the European Community, Spain is the second producer and according to the Annual report from the *Departament d'Agricultura, Ramaderia, Pesca, Alimentació i Medi Natural*, (Gaspa et al., 2011), Catalonia is the autonomous community with the major contribution, accounting for the 40% of total production, followed by Castille and Leon (15%) and Andalusia (9%) (**Figure 1**).

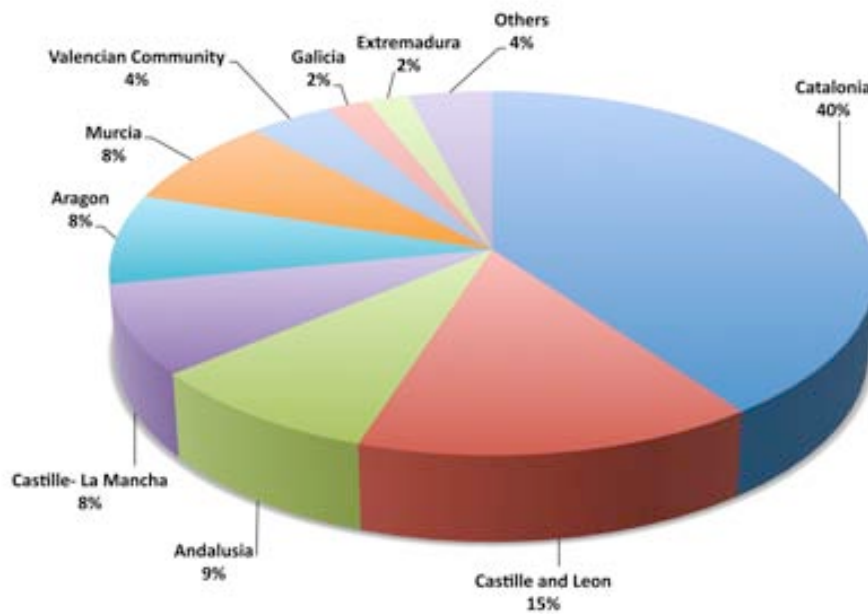


Figure 1. Distribution of pork production among the autonomous communities in Spain. (Source: Ministerio de Agricultura, Alimentación y Medio Ambiente).

Traditionally, pig breeding programs are focused on production efficiency, considering breeding goal traits like feed efficiency, reproductive performance, growth and carcass quality. Consumer preferences for food products refocused industry to the introduction of new breeding objectives, such as meat quality. In the early twenty-century a strong negative selection was applied in commercial breeds against the level of subcutaneous backfat. By a combination of genetics, nutrition and feeding management, a dramatic reduction of fat in pig meat has been produced, decreasing from about 30% of the total carcass in the 1960 to about 15% in an 80 kg carcass of today (Wood et al., 2007). This practice had detrimental consequences on meat quality due to a reduction in the percentage of intramuscular fat (IMF). This is

observed mainly in commercial lean pigs, which show percentages of IMF of 2–5% (Wood et al., 2007).

1.1 - Pork meat quality

Meat quality refers to a broad group of fresh meat processing and sensory characteristics (Cameron, 1993). The term is applicable to fresh meat and to cured products. Therefore, meat quality cannot be measured in a simple and unique manner and different criteria and perceptions are used as: sensorial (eating) quality, nutritional quality, technological quality and hygienic quality (Sellier, 1998). Each of the different meat quality perceptions involves different and related traits. For example, processing meat quality traits consist of firmness, water-holding capacity, cooking loss and other processing yields (Cameron, 1993; Rothschild et al., 2011). On the other hand, sensory traits involve appearance (color and marbling), tenderness, juiciness and flavor of cooked meat (Rothschild et al., 2011). Table 1 shows some of the most frequently considered traits related with the sensorial, nutritional and technological quality of pork meat.

Table 1. Traits related with the sensorial, technological and nutritional quality of pork meat quality (adapted from Sellier, 1998).

Perception	Traits	Attributes of quality
Sensory quality	Texture Flavor Tenderness Juiciness Appearance Color Firmness Marbling	Fresh meat
Technological quality	pH (initial and ultimate) Water-holding capacity Cooking loss Conductivity Muscle glycogen content Characteristics of muscle fiber	Processed pork products
Nutritional quality	Lipid composition Protein Vitamins Minerals Digestibility	

Genetic decomposition of meat quality traits has been extensively studied showing, in general, for sensorial and technological quality traits low or moderate heritabilities with average values into the range 0.10–0.30 (Rothschild et al., 2011; Sellier, 1998). In contrast, for traits determining nutritional quality such as lipid and fatty acid (FA) composition, moderate to high heritabilities ranking from 0.15-0.55 have been reported (Casellas et al., 2010; Ntawubizi et al., 2010). The genetic correlations (r_A) among these traits are variable and in some cases show heterogeneous patterns from tissue to tissue (Casellas et al., 2010; Sellier et al., 2010).

As noted before, technological, nutritional and organoleptic properties of pork meat quality is highly dependent (among other factors) on the lipid content and FA composition (Ntawubizi et al., 2010; Suzuki et al., 2006; Wood et al., 2007; Wood et al., 2008). In addition to being a major energy source, FA are cellular signaling molecules which play relevant roles in various metabolic pathways and have an important impact in human health (FAO, 2010; Okada et al., 2005; Stables et al., 2011; Wakil et al., 2009). Molecular processes controlling FA metabolism are highly interconnected and controlled by environmental and genetic effects. Moreover, according to the nutritional or the eating quality the ideal meat FA composition will differ. For example, PUFAs, mainly ω -3, have been considered beneficial for human health, by reducing serum low-density lipoprotein-C, total cholesterol concentration and modulating immune functions and inflammatory processes (Poudyal et al., 2011; Rudel et al., 1995; Stables et al., 2011). However, from the meat quality point of view, PUFAs have a negative effect on the oxidative stability of muscle, affecting flavor and muscle color (Wood et al., 2008). On the other hand, desirable sensorial characteristics tend to be associated with MUFA and SFA (Carrapiso et al., 2003; Chizzolini et al., 1998; Wood et al., 2008), but a high consumption of SFA is associated with obesity, high plasma cholesterol and cardiovascular diseases in humans (Chizzolini et al., 1999; Katan et al., 1994). Therefore, elucidating the genetic basis of FA composition is an active research area that could support technological development towards improvement of pork meat quality and also increased knowledge of FA metabolism underpinning metabolic human diseases.

1.2 - Pig genomics

The porcine genome has a total of 18 autosomes plus the X/Y sex chromosome pair; it is similar in size, complexity and chromosomal organization to the human genome. Research initiatives to characterize and mapping the pig genome started in the late 1980s under the EU-funded Pig Gene Mapping Project (PiGMaP) (Groenen et al., 2012; Haley et al., 1990). The initial goal of the PiGMaP project was focused on the development of genetic markers and the establishment of a genetic linkage and cytogenetic map (Archibald et al., 1995; Coppieters et al., 1995; Davies et al., 1994; Echard et al., 1992; Groenen et al., 1995). These results facilitated the identification of hundreds of new genetic markers that were integrated into the pig map, allowing the implementation of linkage analysis for the identification of quantitative trait *loci* (QTL) and candidate genes associated with complex traits and also the genetic characterization of pig populations (Andersson et al., 1994; Clop et al., 2003b; Perez-Enciso et al., 2000).

The establishment of the Swine Genome Sequencing Consortium (SGSC) in 2003 constitutes the starting point in the pig genome sequence project. The genomic DNA of a Duroc female pig was obtained at 4-fold coverage from bacterial artificial chromosome (BAC) derived sequence data (Archibald et al., 2010). Then, the genome assembly Sscrofa9 was released in September 2009 and later revised to Sscrofa10 by the addition of whole genome shotgun sequence data (Archibald et al., 2010). The availability of pig genome sequence together with the continuous improvement of the high-throughput sequencing technologies, have generate new opportunities to explore the genetic architecture of complex traits and to study the variability of populations by the discovery of a range of polymorphisms that can be used as genetic markers.

For example, new and valuable tools have been development such as the Porcine 60K SNP Beadchip (Illumina). This chip has been successfully employed in a variety of studies in pigs such as association and linkage analysis (Corominas et al., 2013b; Duijvesteijn et al., 2010; Fernandez et al., 2012; Gregersen et al., 2012; Grindflek et al., 2011; Ramayo-Caldas et al., 2012b), population genetics (Burgos-Paz et al., 2013; Manunza et al., 2013) and the identification of structural variants in pigs, like Copy Number Variants (CNV) (Chen et al., 2011; Chen et al., 2012; Ramayo-Caldas et al., 2010; Wang et al., 2012b).

CNVs constitute an important source of genetic variation and have been associated with both, mendelian diseases and complex genetic disorders in humans, such as developmental delay and intellectual disabilities (Glancy et al., 2009; Ravnan et al., 2006; Sharp et al., 2008), bipolar disorder (Clayton-Smith et al., 2010), body mass index (Sha et al., 2009), schizophrenia (Tam et al., 2009) and Crohn's disease (Schaschl et al., 2009). In domestic animals, some evidences showing the causative effects of CNVs on phenotypic variations have been reported. For instance, a CNV in intron 1 of *SOX5* causing the pea-comb phenotype in chickens (Wright et al., 2009), duplication of *FGF3*, *FGF4*, *FGF19* and *ORAOV1* resulting in hair ridge and predisposition to dermoid sinus in Ridgeback dogs (Salmon Hillbertz et al., 2007), respectively. The wrinkled skin phenotype and the familial Shar-Pei fever disorder characteristic of the Shar-Pei dog are caused by a 16.1-kb duplication near the *HAS2* gene (Olsson et al., 2011) and a 4.6-kb intronic duplication in *STX17* for hair greying and melanoma in horses (Rosengren Pielberg et al., 2008). In pigs, the *KIT* gene was the first example of CNV associated with different phenotypes. The dominant white coat phenotype in pigs is caused by *KIT* gene duplication or triplication and a splice mutation in one of the *KIT* gene copies (Marklund et al., 1998; Moller et al., 1996).

Recently, thanks to the coordinate effort of the SGSC a high quality draft of the pig genome sequence (Sscrofa10.2) have been published (Groenen et al., 2012). Table 2 summarizes some of the assembly and annotation statistics deposited in the Ensembl (release 67) database (Flicek et al., 2011). The assembly Sscrofa10.2 comprise 2.60 gigabases (Gb) assigned to chromosomes with a further 212 megabases (Mb) in unplaced scaffolds (Groenen et al., 2012). The analysis of the pig genome revealed similar evolution rates in pigs and other mammals with an intermediate level of purifying selection pressure in the pig compared with humans and mice (Groenen et al., 2012). The authors also reported that pigs have the largest repertoire of functional olfactory receptor genes and identified 251 putative selective sweeps representing approximate the 1% of the pig genome. Finally, new evidences supporting the usefulness of pig as biomedical model were reported, a comparison among the predicted porcine proteins and their human orthologues showed 112 positions in which pigs have the same amino acid that is implicated in human diseases.

Table 2. Assembly and annotation statistics corresponding to the Sscrofa10.2 genome assembly adapted from (Groenen et al., 2012).

Assembly	Placed	Unplaced	Annotation
Total length (bp)	2.6×10^9	2.1×10^8	21,640 protein-coding genes
Ungapped length (bp)	2.3×10^8	1.9×10^8	380 pseudogenes
Scaffolds	5,343	4,562	2,965 ncRNAs
Contigs	73,524	168,358	197,675 gene exons

1.2.1 - Transcriptome analysis using RNA-Seq

In pigs, several transcriptome architecture descriptions have been performed using the high-density oligonucleotide microarray chips (GeneChip Porcine) from Affymetrix (Santa Clara, CA) which contain a total of 23,937 probe sets, representing 20,201 *Sus scrofa* genes (Fernandez-Rodriguez et al., 2011; Ferraz et al., 2008; Fontanesi et al., 2011; Perez-Enciso et al., 2009; Perez-Montarelo et al., 2012; Ponsuksili et al., 2008). Recently, a new Snowball array have been produced (Freeman et al., 2012). Compared with Affymetrix array, Snowball is more comprehensive in its gene coverage with a total of 47,485 probe sets with a mean coverage of 22 probes/transcript (Freeman et al., 2012).

The massive parallel sequencing methodologies applied to RNA (RNA-Seq) provided new strategies for both transcriptome characterization and gene expression profiling. Recent studies in pigs have employed RNA-Seq as tool to explore the transcriptome (Chen et al., 2011; Esteve-Codina et al., 2011; Jung et al., 2012; Ramayo-Caldas et al., 2012a). In terms of expression levels, this technique overcomes the hybridization-based approaches, e.g., microarrays, due to its higher dynamic range, specificity and sensibility (Marioni et al., 2008; Mortazavi et al., 2008; Wang et al., 2009). Regarding terms of transcriptome architecture, RNA-Seq provides a most complete picture of the transcriptome, allowing the characterization of alternative splicing, variation in the usage of promoters and polyadenilation sites, single nucleotide variants (SNVs), transposable elements and non-coding RNAs. Another important advantage is that RNA-Seq is not limited to detecting transcripts that correspond to existing annotated genes, which is attractive mainly for incomplete annotated genomes or non-model organisms (Davey et al., 2011; Garber et al., 2011; Mortazavi et al., 2008; Wang et al., 2009).

1.3 - Mapping complex traits in pigs

A variety of molecular tools, experimental designs and statistical approaches have been employed to identify QTLs and candidate genes. Among the strategies most often employed in livestock species there are: the candidate gene approach, QTL mapping, genome wide association analysis (GWAS) and, more recently, holistic approaches based on systems biology. A brief description of those methodologies, their advantages and principal limitations are discussed below.

1.3.1 - Candidate gene approach

The candidate gene approach assumes that a gene involved in the physiology of the trait could harbor a mutation responsible for the phenotypic variation in that trait. Afterwards, the gene is sequenced to look for variations into the promoter, introns or coding regions. If polymorphisms are found, the choice of SNPs for genotyping and association analysis depends on the linkage disequilibrium (LD) blocks. The premise of these studies is that SNPs under investigation capture information about the underlying genetic variability of the candidate gene. These strategies have three principal problems. First, due to the genome complexity, the causative mutation may lie in a gene or genome region that would not have been regarded *a priori* as an obvious candidate gene for this particular trait. Second, quantitative traits are usually determined by many genes, a large number of genes must be sequenced and association studies must be carried out in a large number of samples. Finally, the selected SNPs in the candidate gene that are being studied could not have a functional effect (Foulkes, 2009; Hayes, 2007).

1.3.2 - QTL mapping and GWAS

In the late 1990s, advances in molecular genetics and statistical approaches have opened new possibilities for dissecting and understanding the genetic basis of complex traits. QTL analysis is a statistical method that links two types of information: phenotypic and genotypic data. The main goal of QTL mapping is to identify chromosome regions associated with variation in phenotypic traits. If a QTL is linked to a marker, individuals with different marker genotypes will have different mean values for the quantitative trait (Lander et al., 1989; Mackay et al., 2009).

In domestic animals, QTL mapping started with the pioneering work of (Andersson et al., 1994) in pigs. Afterwards, many efforts have been done to detect QTL for a huge variety of economically relevant traits in pigs. In general, the results from QTLs studies are collected in pig QTL database (PigQTLdb) (Hu et al., 2013; Hu et al., 2007). According to the Release 19 (Dec 27, 2012), PigQTLdb contains 8,315 QTLs from 355 publications. These QTLs represent 622 different traits that are distributed in five principal classes (Production, Health, Reproduction, Exterior and Meat & Carcass Quality). Reflecting their economical relevance, Meat & Carcass Quality is so far the class showing the major number of deposited QTLs (5,341), followed by Reproduction (782 QTLs) and Health (775 QTLs). **Figure 2** resume the distribution by chromosomes of QTLs related with meat quality traits.

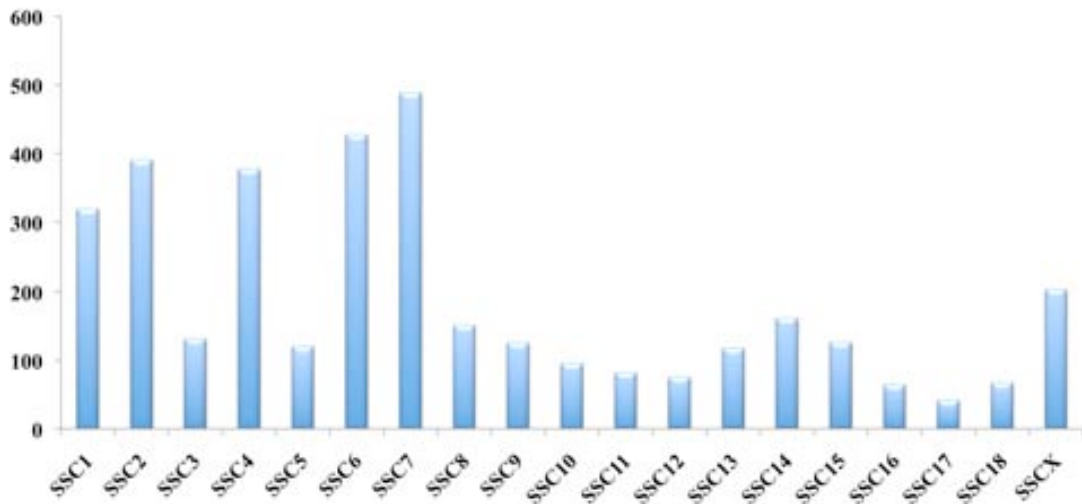


Figure 2. Distribution of reported QTLs for meat quality traits across the pig genome. The number of QTLs corresponds to Release 19 (Dec 27, 2012) of PigQTLdb. X-axis represents the number of QTL and Y-axis the pig chromosome. eQTL mapping results were not included in this figure.

Classical, family-based linkage analysis, with a limited number of microsatellites markers has been used for the detection of QTLs. Due the limited number of recombination events in current animal experimental populations, many microsatellite studies lacked resolution (or power) in QTL detection. Therefore, the linkage often extends to large distances, and the confidence intervals of the identified QTLs are frequently large, from 10 to 20 centiMorgans (cM). This is one important limitation that complicates the fine mapping and the identification of causative mutations (Ledur et al., 2010; Mackay et al., 2009; Meuwissen et al., 2000). Hence, is not surprising that currently only a few causative mutations affecting meat and/or carcass traits have been identified in pigs (Table 3).

Table 3. Functional polymorphisms affecting meat and/or carcass traits in pigs.

Genes	Traits	Reference
Ryanodine receptor 1 (<i>RYR1</i>)	Malignant hyperthermia syndrome and pale soft and exudative meat	Fujii et al., 1991
Protein kinase AMP-activated gamma 3-subunit (<i>PRKAG3</i>)	Glycogen metabolism, Napole yield, pH and color	Milan et al., 2000; Ciobanu et al., 2001
Insulin-like growth factor 2 (<i>IGF2</i>)	Carcass composition, backfat thickness	Van Laere et al., 2003
Melanocortin 4 receptor (<i>MCR4</i>)	Carcass composition, backfat thickness	Kim et al., 2000

Furthermore, most of the QTLs identified tended to be only the ones with large effect. Hence, relatively few, but important QTLs were mapped with microsatellites (Andersson et al., 1994; Clop et al., 2003a; Dekkers, 2004; Goddard et al., 2009; Perez-Enciso et al., 2000). Finally, the practical implementation of QTL strategies in marker assisted selection (MAS) is limited because the linkage between markers and QTL is not sufficiently high to ensure that marker-QTL allele relationship persist across the population (Hayes, 2007). Those previously mentioned limitations of QTL analysis with microsatellites could be partially solved using a more dense set of markers and by exploiting linkage disequilibrium (LD) to map QTLs (Meuwissen et al., 2000). Nowadays, thousands of SNPs covering the whole genome can be genotyped in a number of individuals at reasonable costs, which dramatically increases the resolution achieved from mapping studies (Perez-Enciso et al., 2010). The replacement of microsatellites with high-density SNP arrays was followed by the development of a range of statistical approaches applied to GWAS and/or genomic selection. High-density SNP arrays allow for more QTLs discoveries and facilitate the identification of candidate genes as well as causative mutations

(Goddard et al., 2009; Meuwissen et al., 2000). However, the utilization of LD to map QTL still has some limitations. For instance, the pattern of LD observed in a population depends on its history, mainly the history of its effective population size (N_e) (Hayes et al., 2003; Nordborg et al., 2002; Sved, 1971). Domestication and breed formation processes caused a significant reduction of N_e in livestock; hence some LD patterns extend long distances ($> 1\text{cM}$). This long LD pattern is common for many of the domestic animals but is the opposite of that observed in humans (Goddard et al., 2009). For instance, at short genetic distances humans have similar LD to cattle but almost no LD at long distances (Goddard et al., 2009). Consequently, the SNP density required for a GWAS in domestic animals is lower than in humans. Conversely, it will not be possible to localize QTLs precisely (in within-breed association mapping), making fine mapping more difficult (Goddard et al., 2009; Sutter et al., 2004). Nevertheless, because long-range LD does not apply across breeds, this problem could be overcome by the analysis of multiple breeds (Goddard et al., 2009).

In a GWAS, a statistical association between a trait of interest and genetic markers (typically SNPs from a genome-wide panel) is identified assuming that a significant association can be detected because the SNPs are in LD with the causative mutations for the traits of interest (Foulkes, 2009; Klein et al., 2005; Mackay et al., 2009). In animal production, probably the most used GWAS approach has been a single-trait-single-SNP, which is based on a full animal model and uses pedigree information. In comparison with the classical QTL mapping, GWAS has the advantage of using all recombination events after the mutations occurred and, therefore, increases the precision of the QTL position estimates and reduces their confidence intervals (Meuwissen et al., 2000). In addition, GWAS considers the contribution of the variability within breeds or lines, whereas linkage analysis usually

ignores it and, as a consequence, its power is decreased (Ledur et al., 2010). Moreover, the use of a dense set of markers provides the unique opportunity to simultaneously exploit linkage and LD for QTL fine-mapping (LDLA) (Druet et al., 2009; Meuwissen et al., 2004). This approach has been successfully employed to fine mapping QTLs and reduced the QTL confidence intervals in several species such as dairy cattle, pigs and horses (Corominas et al., 2013b; Dupuis et al., 2011; Karim et al., 2011; Kuehn et al., 2007; Ren et al., 2011; Sartelet et al., 2012).

One important limitation of GWAS is the lack of replication of the results in an independent population, limiting therefore its practical application in breeding programs through marker-assisted selection (MAS). In fact, lessons learned from GWAS in humans revealed that most of the confirmed replication occurs only when a large sample size (thousands of individuals) was used in meta-analyses (Asselbergs et al., 2012; Klein et al., 2012; Saxena et al., 2012; Stranger et al., 2011). On the other hand, the maximum successes in identifying candidate genes using GWAS have been reached for monogenic traits (Georges, 2007; Goddard et al., 2009; Mackay et al., 2009). For complex traits the results are in agreement with the well know genetic architecture of polygenic traits, observing many mutations each one having a small effect. For example, GWAS for milk yield and reproductive traits in cattle have reported several significantly-associated SNPs (Hawken et al., 2012; Hayes et al., 2009; Huang et al., 2010; Jiang et al., 2010; Mai et al., 2009; Sahana et al., 2010). In pigs, the first published GWAS using the genotype information from the Illumina Porcine 60K SNP Beadchip was done by (Duijvesteijn et al., 2010). The authors reported 37 SNPs affecting androsterone level in fat tissue. Then, a large-scale genome-wide association and LDLA mapping reported 28 chromosomal regions related to boar taint and related sex steroids traits (Grindflek et al., 2011). Recently,

several GWAS showing similar patterns have been performed for a variety of economically relevant traits in pigs such as feed efficiency, growth and fatness, meat quality and reproduction (Fontanesi et al., 2012; Luo et al., 2012a; Luo et al., 2012b; Onteru et al., 2011; Sahana et al., 2013; Schneider et al., 2012).

1.3.2 - Gene Network Approach for the study of Complex Traits

Complex phenotypes are determined by a set of inter-related genes and their interaction with environmental factors. Both GWAS and linkage-based methodologies examine the association of a chromosomal region (QTL scan) and/or single genetic variants (GWAS) at time, ignoring therefore their functional interactions. A perturbation in one gene can be propagated through their interactions, and affect other genes and/or transcription factors (TF) which in turn regulates thousands of genes (Califano et al., 2012; Cho et al., 2012). So, due to the complex architecture and synergistic effects among these genes, the holistic effect of a gene network or a pathway is expected to have a larger effect than the sum of small individual effect of each gene (Sun, 2012). On the other hand, an important advantage of working with gene modules rather than individual genes is that it is often easier to predict the function of the module than the function of a single gene (Cho et al., 2012; De Smet et al., 2010).

In recent years, systems biology approach, more specifically network-based approach, has emerged as an alternative and powerful tool to study complex traits. Systems biology is not focused on the components themselves, but on the nature of the links that connect them and the functional states of the networks that result from the assembly of all such links (Palsson, 2006). The basic idea of a biological network is to represent the relationships between biological entities based on their functional

connections (Filkov, 2005; Sun, 2012). Different types of biological networks such as protein-by-protein interaction, metabolic networks and gene regulatory networks exist. Their definitions are often determined by the biological entities under study and the nature of their relationship. For instance, in a protein-by-protein interaction network the nodes represent proteins and the edges a physical and bidirectional interaction among the proteins (De Smet et al., 2010; Sun, 2012). In a gene regulatory network, nodes could be represented as DNA segments (genes, TF, cis-elements and/or trans-elements) and interactions among them may be inferred, for instance, by their co-expression patterns across multiple scenarios or conditions (De Smet et al., 2010; Filkov, 2005). Finally, integrative networks combine the expression data with complementary information such as metabolites or chromatin immunoprecipitation-on-Chip data. They can obtain reliable interactions and a more complete picture of the biological process under consideration. (De Smet et al., 2010). Several algorithms, implemented in multiple programs have been developed to deal with these heterogeneous data. Among the most popular algorithms there are Partial Correlations, Ordinary Differential Equations, Supervised Machine-Learning Approach, Gaussian and Bayesian models (Abdolreza Mohammadi et al., 2012; Aravena et al., 2013; Bansal et al., 2006; Langfelder et al., 2008; Mordelet et al., 2013; Reverter et al., 2008; Zhang et al., 2010; Zhu et al., 2004; Zhu et al., 2012).

Recently, many studies have employed these approaches to facilitate the identification of the causal genes and the characterization of the underlying molecular networks that drive the traits of interest, mostly in humans, mice and yeast (Emilsson et al., 2008; Schadt et al., 2005; Zhu et al., 2007; Zhu et al., 2008). In livestock, most of the studies have been focused in the inference and implementation of gene weighted co-expression analysis (Badaoui et al., 2013; Gu et al., 2011; Harhay et al.,

2010; Hudson et al., 2012; Kogelman et al., 2011; Stanley et al., 2013; Sun et al., 2012; Yang et al., 2011), protein-by-protein interaction networks (Jiang et al., 2011; Lim et al., 2011; Wang et al., 2012a) and gene networks derived from co-association analysis (Fortes et al., 2010; Fortes et al., 2011; Fortes et al., 2012). In pigs, few studies have been implemented (Badaoui et al., 2013; Freeman et al., 2012; Perez-Montarelo et al. 2012; Wang et al., 2012a; Yang et al., 2011). Moreover, as far as we know, despite its economic relevance, no study has been published on the genetic dissection of meat quality traits.

CHAPTER II



OBJECTIVES


CHAPTER 2. OBJECTIVES

The broad goal of this thesis was the identification of genetic markers, genes and gene networks associated with the profile of intramuscular fatty acid composition in pigs. This work was done under the framework of the IbmAP project funded by the projects AGL2008-04818-C03/GAN (MICINN) and AGL2011-29821-C02 (MINECO).

More specifically, the objectives were:

- 1- To identify Copy Number Variant Regions in the pig genome.
- 2- To pinpoint by genome-wide association analysis chromosomal regions and positional candidate genes associated with the intramuscular fatty acid profile and indices of fatty acid metabolism.
- 3- To identify differentially-expressed genes in the liver of Iberian x Landrace crossbred pigs showing extreme phenotypes for intramuscular fatty acid composition using RNA-Seq.
- 4- To infer a gene network for intramuscular fatty acid composition in pigs employing a systems biology approach based on a SNP-to-SNP co-association.

CHAPTER III
GENERAL
DISCUSSION



“There is no privileged level of causality in biological systems”

- Denis Nobel

CHAPTER 3. GENERAL DISCUSSION

The molecular processes controlling FA composition and metabolism are highly interconnected and not fully understood. Elucidating these molecular processes will aid the technological development towards the improvement of pork meat quality and will increase our knowledge of metabolic diseases in humans. This thesis deals with FA metabolism in pigs, analyzed from different perspectives. We have described for the first time a Copy Number Variant Regions (CNVR) map in swine based on whole genome SNP genotyping chips, some of them may have an effect on FA (eg, CNVR6, CNVR11, CNVR34 and CNVR36). In addition, we have employed a different but complementary approach, a GWAS in a Iberian x Landrace backcross across 32 traits related to IMF FA composition to identify 813 associated SNPs distributed in 43 chromosomal intervals across the pig genome (Ramayo-Caldas et al., 2012b). Positional concordance among 23 (53.5%) of these regions and previously reported QTL for FA composition and/or other lipid traits were found. These common genomic regions for different traits were primarily found in SSC4, SSC8 and SSC16, suggesting pleiotropic effects of some *loci* on FA composition. It is noteworthy that six of the associated genes identified by GWAS (*APOB*, *CYP7A1*, *APOA2*, *THBS1*, *THEM5*, *ME3*) were later reported as differentially-expressed (DE) in the hepatic transcriptome analysis of pigs with divergent phenotypes for IMF FA composition (Ramayo-Caldas et al., 2012a).

In addition, other 49 protein-coding genes (55 genes in total) were identified as DE in the RNA-Seq study in liver. Functional analysis revealed that these genes belong to biological functions and canonical pathways closely related to lipid and FA metabolism processes. Importantly, three interconnected gene-networks were identified, suggesting that the DE genes detected in this study are functionally linked and play an important, coordinated role in lipid metabolism. The complexity and high interconnectivity of the molecular processes controlling FA metabolism identified by our GWAS (Ramayo-Caldas et al., 2012b) and RNA-Seq (Ramayo-Caldas et al., 2012a) studies, motivated us to use a holistic gene network approach based on SNP-to-SNP co-association analysis (Fortes et al., 2010). For the first time, supporting evidence for co-association network predictions were confirmed by gene co-expression analysis carried out in two relevant lipogenic tissues: adipose and liver. The analysis of the topological properties of both the co-association and co-expression predicted gene networks, allowed the identification of key TF, candidate genes and metabolic pathways that are likely being determining meat quality and FA composition, as well as controlling energy homeostasis in pigs.

3.1 - Detecting CNVR in the pig Genome

Advances in whole-genome technologies, including array comparative genomic hybridization (CGH), single nucleotide polymorphism (SNP) microarrays, and genome sequencing, have enabled the discovery and characterization of CNVs (Teo et al., 2012; Xi et al., 2010). Here, we analyzed genotyping data from the Porcine SNP60 BeadChip and a total of 49 autosomal CNVRs were reported. Mendelian inheritance of CNVRs was confirmed and a segregation analysis of 560 pigs with different geographical origins showed a large between-breed variation. In

addition, five of the seven-selected CNVs were successfully validated by qPCR, representing a false discovery rate (FDR) of 29%. Similar patterns of FDR have reported in subsequent studies, for example (Wang et al., 2012b) report a FDR ~33%, whereas a FDR ~15% was reported by (Chen et al., 2012).

Remarkably, 26 of the CNVRs identified in our research (53%) have been also reported in later studies (Table 4). A detailed examination of these 26 independent confirmed CNVRs revealed that three (CNVR4, CNVR34, CNVR45) have been reported using three different whole-genome technologies (CGH, SNP arrays and NGS). Eight were reported by at least two technologies (CNVR5, CNVR7, CNVR11, CNVR15, CNVR21, CNVR30, CNVR31 and CNVR36) and the rest (15) were identified using the Porcine 60K SNP Beadchip (Table 4). Moreover, a total of four CNVRs were reported by three different studies (CNVR5, CNVR34, CNVR36, CNVR45), nine were replicated by at least two studies and 13 were reported in one independent dataset.(Table 4). Among the replicated CNVRs, CNVR34 and CNVR45 were supported by three different whole-genome technologies and also reported in three independent studies. CNVR45 contain the KIT gene, which is a well-characterized and functionally important CNV in pigs (Marklund et al., 1998; Moller et al., 1996; Pielberg et al., 2002). As expected, the five CNVRs validated by qPCR in our study have been later confirmed in independent studies. For instance, CNVR15 was identified by (Chen et al., 2012; Rubin et al., 2012) using the Porcine 60K SNP Beadchip and NGS data, respectively, and CNVR36 was identified using the same technologies by (Melo, 2012; Rubin et al., 2012; Wang et al., 2012b).

Table 4. Summary of independent confirmations of 26 CNVRs

CNVR	Reported by
CNVR1	(Chen et al., 2012; Melo, 2012)
CNVR2	(Chen et al., 2012)
CNVR4	(Li et al., 2012; Rubin et al., 2012)
CNVR5	(Chen et al., 2012; Rubin et al., 2012; Wang et al., 2012b)
CNVR6	(Wang et al., 2012b)
CNVR7	(Li et al., 2012; Wang et al., 2012b)
CNVR11	(Chen et al., 2012; Rubin et al., 2012)
CNVR13	(Wang et al., 2012b)
CNVR15	(Chen et al., 2012; Rubin et al., 2012)
CNVR17	(Melo, 2012)
CNVR21	(Li et al., 2012; Wang et al., 2012b)
CNVR22	(Wang et al., 2012b)
CNVR23	(Chen et al., 2012)
CNVR26	(Chen et al., 2012)
CNVR30	(Li et al., 2012; Wang et al., 2012b)
CNVR31	(Li et al., 2012; Wang et al., 2012b)
CNVR32	(Wang et al., 2012b)
CNVR33	(Wang et al., 2012b)
CNVR34	(Li et al., 2012; Rubin et al., 2012; Wang et al., 2012b)
CNVR35	(Melo, 2012; Wang et al., 2012b)
CNVR36	(Melo, 2012; Rubin et al., 2012; Wang et al., 2012b)
CNVR38	(Wang et al., 2012b)
CNVR41	(Wang et al., 2012b)
CNVR42	(Melo, 2012)
CNVR43	(Wang et al., 2012b)
CNVR45	(Chen et al., 2012; Rubin et al., 2012; Wang et al., 2012b)

It should be noted that gene *CYP2C49*, located within CNVR36 was also identified as DE in our RNA-Seq study (Ramayo-Caldas et al., 2012a). In order to evaluate whether the observed differences in gene expression were influenced by the number of copies of the gene. A qPCR assay to determine the number of copies of the *CYP2C49* gene was developed. However, no correlation between the number of copies and gene expression was observed. A similar situation has been reported in rats and mice in which only a fraction of genes showed a strong correlation between the number of copies and the expression level (Guryev et al., 2008; Henrichsen et al., 2009b). Dosage compensation, lack of regulatory elements in the duplicated copies or differences in the chromatin environment might partially explain these observations (Henrichsen et al., 2009a; Henrichsen et al., 2009b). In addition to CNVR36, and according their biological function, there were other regions that may have an effect on FA. For instance, the Acetyl-CoA Acyltransferase 2 gene (*ACAA2*) mapped within the CNVR6, this gene belongs to the FA metabolism pathway and catalyzes the last step of the mitochondrial fatty acid beta-oxidation. CNVR11 contains the very low-density lipoprotein receptor gene (*VLDLR*), which plays important roles in VLDL-triglyceride metabolism. Finally, Ryanodine Receptor 2 gene (*RYR2*) is located within CNVR34. It should be noted that a role of both *VLDLR* and *RYR2* genes controlling the plasma level of lipids in human has been recently reported (Asselbergs et al., 2012). Moreover, positional concordance among these CNVRs and fat-related QTL deposited in the Pig QTL Database were observed (Table 5).

Table 5. Positional concordance among candidate genes for FA metabolism located within CNVRs and QTLs deposited in the pig QTL database for fatness related traits.

CNVR	Candidate gene	Trait	QTL Id
CNVR6	<i>ACAA2</i>	Average backfat thickness	5670
		Fat to meat ratio	5677
		Marbling	2930
		Cholesterol level	17711
		HDL/LDL ratio	17712
		LDL cholesterol	17713
CNVR11	<i>VLDLR</i>	Linolenic acid percentage	686
		Average backfat thickness	4016
		Backfat above muscle dorsi	5673
		Shoulder subcutaneous fat thickness	5671
CNVR34	<i>RYR2</i>	Stearic acid percentage	15870
		Linoleic acid percentage	15871
		Backfat above muscle dorsi	5991
		Average backfat thickness	37
CNVR36	<i>CYP2C49</i>	Stearic acid percentage	15870
		Average backfat thickness	1161
		External fat on ham	1167
		Adipocyte diameter	12850

Although we failed to validate by qPCR the CNVR1, changes in the number of copies overlapping this region have been reported in two independent studies (Chen et al., 2012; Melo, 2012). Several factors may account for the discrepancy in CNVR prediction between the *in silico* analysis and the qPCR method. First, the low probe density of the Porcine 60K SNP Beadchip makes it difficult to establish the true boundaries of CNVRs. Therefore, it cannot be ruled out that the primers used to validate the CNVRs by qPCR may have been designed outside the structural polymorphic region. Second, polymorphisms such as SNPs and indels may influence the hybridization of the qPCR primers, changing the relative quantification (RQ) values for some animals. Finally, the true CNVR boundaries may be also polymorphic between the analyzed animals. The CGH or SNP arrays have important limitations to detect

small CNVs, even for arrays containing more than 1 million probes the resolution is limited to 10-20 kb (Carter, 2007; Cooper et al., 2008).

Mapping accurately the boundaries of CNVs and improving the level of resolution in CNV calling are essential to estimate their impact on complex phenotypes. Both parameters could be substantially improved by using high-throughput sequencing at a high coverage (Alkan et al., 2009; Teo et al., 2012; Xi et al., 2010). However, lessons learned from humans revealed some important limitations and challenges for CNV detection using NGS data (Teo et al., 2012; Xi et al., 2010). For example, high-quality assembled reference genomes are required (which is so far to be a reality in pigs). Other important limitations affecting the performance of algorithms for CNV detection are: 1) an elevated sequencing error rate of NGS platforms, producing a considerable loss of reads during alignment (Dohm et al., 2008; Xi et al., 2010); 2) the lack of enough sequence depth to obtain reasonable values of sensitivity and specificity (Xi et al., 2010); 3) given the currently short read length of NGS sequences many reads do not map to unique positions in the genome and therefore structural variants in these regions are frequently not effectively being identified; 4) the insert size of the sequencing libraries is also an important parameter for structural variant detection, at the same sequencing coverage, smaller insert sizes will be more accurate to localizing breakpoints and to detected smaller events than larger insert sizes (Teo et al., 2012, Bashir, 2008; Xi et al., 2010).

3.2 - From GWAS to Gene-Networks through expression and co-expression analysis

Previous studies in the IBMAP cross reported QTLs for the FA composition in the backfat of an F2 cross animals (Pérez-Enciso et al., 2000; Clop et al., 2003). Although full concordance between the chromosomal regions identified in our GWAS and those QTLs previously reported were not expected, some interesting positional concordance was observed mainly in SSC4 and SSC8, suggesting QTLs with pleiotropic effects. It should be noted that recently (Muñoz M. et al 2013, submitted) have confirmed our initial hypothesis of pleiotropic effects of those QTLs determining the FA composition in both, muscle and backfat tissues. Reported QTLs for FA composition in the backfat of F2 animals were identified by linkage analyses using few microsatellite markers, therefore the QTL confidence interval extended to large distances. For example, the confidence interval corresponding to the QTL reported by Clop et al 2003 in SSC8 for palmitoleic acid was approximately 17 cM. Initially, the *MTTP* gene was evaluated as positional candidate gene for this QTL and a mutation in the lipid transfer region of the MTTP protein (p.Phe840Leu) was associated with FA composition of porcine fat and with the MTTP lipid transfer activity (Estelle et al., 2009). Recently, a combination of GWAS and LDLA analysis allowed us to reduce the confidence interval for this QTL for palmitic acid at position 117,824,360–119,887,525 bp, whereas for palmitoleic acid at position 117,824,360–119,727,822 bp. These positions maximized the QTL peak at approximately 10 Mb from the *MTTP* gene in the region where the *ELOVL6* gene was located. A strongly associated polymorphism with the percentage of palmitic and palmitoleic acids in muscle and adipose tissues was identified in the promoter region of *ELOVL6* gene, *ELOVL6*:c.-533C>T (Corominas et al., 2013b). Interestingly, both *MTTP* and *ELOVL6* genes were co-associated in the AWM gene-network and also a moderate significant co-expression between them was

observed in the liver network ($r=0.53$, $P < 0.01$). Supporting the observed co-expression and the possibility that *MTTP* and *ELOVL6* share a common regulatory mechanism, a significant expression QTL (eQTL) regulating the liver expression of both genes has been recently identified (María Ballester personal communication). Therefore, a genetic effect of this QTL on both *ELOVL6* and *MTTP* genes could not be ruled out and needs further elucidation.

A comparison of the list of genes identified by GWAS with the 55 genes reported as DE in the liver transcriptome showed an overlapping of six genes: *APOB*, *CYP7A1*, *APOA2*, *THBS1*, *THEM5* and *ME3*. The biological relevance of *APOB*, *CYP7A1* and *THBS1* controlling the plasma level of lipids in humans has been recently reported in two meta-analyses (Asselbergs et al., 2012; Teslovich et al., 2010). Moreover, *APOB* and *THEM5* have been also identified as DE in an analysis of the adipose tissue transcriptome (Corominas et al., 2013a submitted). Both liver and adipose tissues play a crucial role in regulating pig lipid metabolism (Nafikov et al., 2007; O'Hea et al., 1969). Therefore, these genes (*APOB*, *CYP7A1*, *THBS1* and *THEM5*) may play an important role in the FA metabolism processes in pigs in both liver and adipose tissues. Finally, 32 of the genes identified by GWAS were later reported as DE in adipose tissue (Corominas et al., 2013a submitted) and 18 in the muscle transcriptome of pigs with extreme phenotypes for IMF FA composition (Puig-Oliveras et al., 2013 submitted).

To gain insight into the liver processes that are different between the two groups of RNA-Seq animals, the list of DE genes was explored using the core analysis function included in Ingenuity Pathways Analysis (IPA). The top biological functions identified by IPA included categories related to a wide variety of physiological and biological events, such as lipid metabolism, small molecule biochemistry, molecular transport,

drug metabolism and energy production. A specific examination of the lipid metabolism IPA molecular and cellular function revealed that most of the transcripts related to lipid metabolism were up-regulated in the High group (H) compared to Low group (L). Remarkably, genes that play a crucial role in lipoprotein synthesis (*APOB*), cholesterol metabolism (*ABCG8*, *CYP2C9*, *CYP2C19*, *CYP4A11* and *CYP7A1*), oxidation of lipids and palmitic fatty acids (*MTMR7*), and induction of lipogenic gene transcription (*LPINI*) were up-regulated in H group in contrast to L group. On the contrary, genes involved in the accumulation of triacylglycerol (*AQP7*), uptake of lipids and myristic acid (*THBS1*), and fatty acids biosynthesis (*ME3*) were down-regulated in H group. In concordance with the initial phenotypic classification, pathway analysis inferred that linolenic and arachidonic acid metabolism was altered between extreme animals. Finally, three interconnected gene-networks were identified, suggesting that the DE genes detected in this study are linked and play an important role in lipid metabolism.

Finally, in order to identify key regulators of FA metabolism, we employed a previously described systems biology approach termed Association Weight Matrix (Fortes et al., 2010). The gene-network built from the SNP-to-SNP co-association provides a clear evidence for the complexity of molecular processes controlling FA metabolism. A total of 1,096 nodes linked by 111,198 significant interactions were observed. We explored the network and selected the TF trio that spanned most of the network topology with a minimum redundancy (*NCOA2*, *FHL2* and *EP300*). The promoter region of these genes contain transcription factors binding sites for some well known TF that regulate lipid and carbohydrate metabolism such as: *SREBP-1*, *PPARG*, *PPAR- α* , *HNF1A*, *HNF4- α* , *ER- α* and *GR- α* . Remarkably, according to the String database (Snel et al., 2000; Szklarczyk et al., 2011) (<http://string-db.org/>), experimental data confirmed a functional interaction between these TF (*NCOA2*, *EP300*, *FHL2*) and the TF considered master regulators of lipid and carbohydrate metabolism (**Figure 3**).

In agreement with this observation, a significant co-expression in the liver network among *NCOA2* and *EP300* and some of this TF was observed. For instance, *NCOA2* was significantly co-expressed with *PPAR-α* ($r = 0.39$, $P < 0.01$), *HNF1A* ($r = 0.56$, $P < 0.001$) and *HNF4-α* ($r = 0.36$, $P < 0.01$). Finally, *EP300* was co-expressed with *HNF1A* ($r = 0.64$, $P < 0.001$).

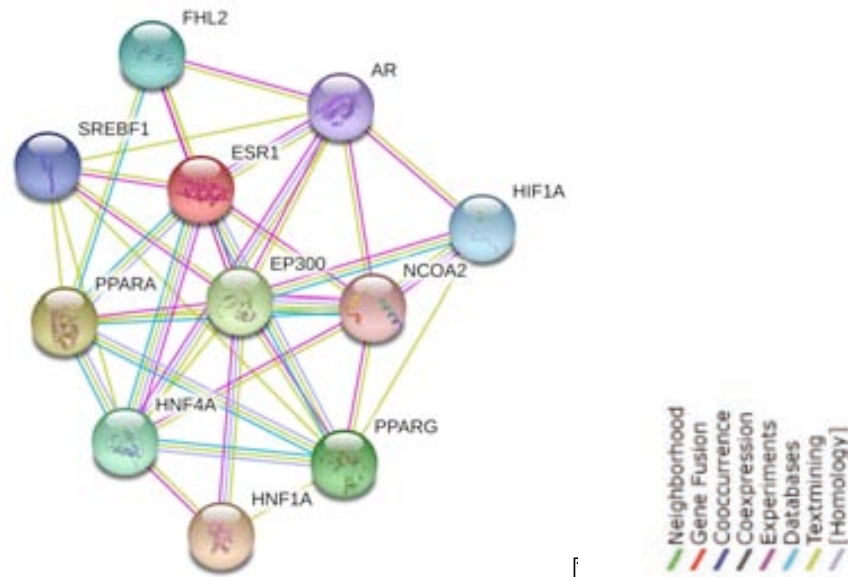


Figure 3. Evidence of functional interactions among the three TF (*NCOA2*, *EP300*, *FHL2*) and TF considered as master regulators of lipid and carbohydrate metabolism.

Remarkably, 39 of the 730 predicted target genes have been recently reported in two large-scale meta-analyses for plasma lipids in humans (Asselbergs et al., 2012; Teslovich et al., 2010). It should be noted that many of these genes, including our TF trio and other FA relevant genes, would have been missed by traditional single-trait GWAS due to the lack of an acceptably significant association level. Enrichment analysis revealed an overrepresentation for GO terms such as “Cellular component organization” ($P = 4.74 \times 10^{-10}$, $FDR = 1.82 \times 10^{-8}$), “Anatomical structure morphogenesis” ($P = 3.49 \times 10^{-6}$, $FDR = 3.88 \times 10^{-5}$), “Kinase activity” ($P = 1.02 \times 10^{-5}$,

FDR = 9.04×10^{-5}), “Lipid metabolic process” ($P = 5.71 \times 10^{-4}$, FDR = 2.05×10^{-3}), “Enzyme regulator activity” ($P = 2.21 \times 10^{-3}$, FDR = 3.78×10^{-3}) and “Lipid binding” ($P = 1.34 \times 10^{-2}$, FDR = 3.06×10^{-2}). A detailed examination of genes underpinning the enrichment for “Lipid metabolic process” revealed that most of them belong to three functional categories: “Metabolism of membrane lipid derivative” (32 genes, $P = 9.67 \times 10^{-6}$), “Synthesis of lipid” (45 genes, $P = 3.69 \times 10^{-4}$) and “Fatty acid metabolism” (36 genes, $P = 1.39 \times 10^{-3}$). Pathway analysis showed that the most representative pathways related to the 730 AWM-predicted target genes were “HIF-1 signaling pathway (hsa04066)”, “Acute myeloid leukemia (hsa05221)”, “Colorectal cancer (hsa05210)”, “Renal cell carcinoma (hsa05211)” and “Type II diabetes mellitus (hsa04930)”.

In order to validate the *in silico* predicted AWM interactions, co-expression analysis in liver and adipose tissue was performed. In concordance with recent results in yeast (Wang et al., 2013), we observed that interacting *loci* could jointly regulate the co-expression patterns of pairs of genes. However, it is worth noting that the magnitude of this validation varied in a tissue-specific manner and only the 35.7% of the interactions validated in the adipose tissue were also validated in the liver co-expression analysis. Interestingly, these co-associated and co-expressed genes belong to biological processes related to lipid metabolism including: “Negative Regulation of Fat Cell Differentiation (*INSIG1*, *TCF7L2*, *ZFPM2*)”, “Androgen Receptor Signalling Pathway (*EP300*, *FHL2*, *NCOA2*)”, “Response to Hormone Stimulus (*ABCC5*, *ANGPT1*, *FABP3*, *EP300*, *SORT1*, *FHL2*)” and “Lipid Metabolic Process (*PBX1*, *INSIG1*, *FABP3*, *FDFT1*, *PIP5K1A*, *MAX*, *AASDH*)”.

Overall, our GWAS and network predictions, supported by bibliography and co-expression analysis in liver and adipose tissues, suggest a co-operative role for the three TF (*NCOA2*, *EP300*, *FHL2*) in the transcriptional regulation of IMF FA composition and the control of energy homeostasis in pigs. Despite the prominence of genes and TF mentioned above, our results, as well as the literature, indicate that there is not one single TF/gene, which is more important than all others. On the contrary, our results show numerous interacting genes and TF, organized in pathways without a clear privileged level of causality control FA metabolism. Future studies targeting these genes, their pathways and interactions will continue to expand our knowledge of molecular control of FA metabolism and it might lead to discovery of functional mutations, unfolding new strategies for improve pork meat quality.

3.2 - Future directions

In this thesis, hundreds of genes were shown to be associated with IMF FA composition and thousands of interactions between the associated genes were predicted. We have also demonstrated that the systems biology approach increases the power to identify candidate genes beyond the one-dimensional approach for identifying genes affecting single traits. Furthermore, epistatic interactions between TF and their target genes are likely to contribute to the complex inheritance of FA composition and related polygenic traits (lipid metabolism and energy homeostasis). However, as has been recently demonstrated by the ENCODE project, most biologically meaningful variation is likely to be regulatory, suggesting that most of the causative SNPs will be non-coding (Maher, 2012). Hence, a full catalog of different sources of genetic variation, including SNPs and structural variants (CNVs, inversions, translocations and indels), is required and will be essential to establish appropriate strategies in animal breeding programs.

New high-throughput technologies have revolutionized biological research providing vast amounts of different sources of data. The real challenge for the research communities is to provide functional links to bridge the knowledge gap between these heterogeneous data and complex phenotypes. In this direction, integrative approaches that combine multiple sources of information could provide a more complete picture of the biological processes under consideration. For instance, the implementation of systems biology approaches to analyze metagenomic data and to explore the relationship between a host and its microbes could provide valuable and relevant information of the interactions between parts of complex biological systems (Thiele et al., 2012).

In the context of pork meat quality, an inverse relationship exists between nutritional value and eating quality of pork meat. For instance, SFA are associated with obesity, high plasma cholesterol and cardiovascular disease, but desirable sensorial characteristics of meat tend to be associated with MUFA and SFA. Conversely, PUFA mainly ω -3 have been considered beneficial for human health, however has a negative effect on the oxidative stability of muscle, which, in turn, affects flavor and muscle color (Katan et al., 1994; Wood et al., 2008). Hence, establish a selection criterion to improve meat quality from both the sensorial and nutritional point of view is a complex matter. In our opinion, a holistic and multidisciplinary approach integrating disciplines such are metagenomics and/or nutrigenomics may be required to improve the pork meat quality from both points of view. For example, different to fish, mammals lack the Δ 12 and Δ 15-desaturases required to synthesize FA of n-6 and n-3 series. Hence, PUFAs linoleic (C18:2 n-6) and α -linoleic (C18:3 n-3) acids must be provided by the diet, and their tissue concentration respond to dietary changes. Therefore, swine meat enriched with ω -3 FA can be achieved by feeding with commercial diets supplemented with this PUFA (Guillevic et al., 2009; Pascual et al., 2007; Sioutis et al., 2008) and possibly by selective breeding programs. In fact, our RNA-Seq study suggest that pigs from H group have a higher absorption or PUFA and/or differential metabolism favoring the accumulation in muscle of these FA (differences in elongation, desaturation and oxidation of long-chain ω -3 and ω -6 FA). Previous studies have reported that, in general, fatter pigs show higher proportion of SFA and MUFA, but less PUFA than lean pigs (Estévez et al., 2003; Serra et al., 1998). It is also well known that Iberian pigs have, in backfat and LD muscle, higher percentages of palmitic acid, oleic acid, SFA and MUFA, and lower concentrations of linoleic and α -linoleic acids than commercial breeds. Since all animals in our RNA-Seq experiments were raised and fed under the same standard management conditions, we hypothesized that differences between H and

L groups were caused by the segregation within the analyzed animals of Iberian and Landrace alleles. Despite their physiological differences in terms of ω -3 PUFA metabolism, a similar observation suggesting a genetic basis for ω -3 PUFA level in Atlantic salmon have been reported (Schlechtriem et al., 2007). The author's demonstrate that irrespective of diet, variation exists in the content of ω -3, and individual animals can display an enhanced ability to maintain high levels of ω -3 in their flesh. However, later studies of the same group were not able to link the inter-family differences in ω -3 content to effects on lipid metabolism, including transcriptional modulation of the LC-PUFA biosynthesis pathway (Morais et al., 2012). Conversely, we think that genetic variation was indeed a significant factor affecting meat IMF PUFA content and composition in pigs.

There is scientific evidence that dietary FA exert many of their biological effects through the modulation of gene transcription (Afman et al., 2012; Poudyal et al., 2011; Schmitz et al., 2008). For instance, elevate PUFA bioavailability can be considered as an important factor in the inhibition of the de novo saturated FA synthesis. In agreement with the phenotypic differences in terms of FA composition, we observed that pigs from H group increased the expression of gene sets regulated by PPAR- α (*APOA2*, *CYP2C9*, *CYP2C19*) and RXR (*ABCG8*, *CYP7A1*). This is also in line with the lower IMF content of H group animals in comparison with L animals, and the lower proportion of SFA and MUFA in the IMF. In concordance with our results, **Figure 4** shows as ω -3 acids can act as ligands/modulators of TF like PPAR- α and RXR, down-regulating genes involved in lipid synthesis and stimulating fatty acid degradation (Poudyal et al., 2011; Schmitz et al., 2008).

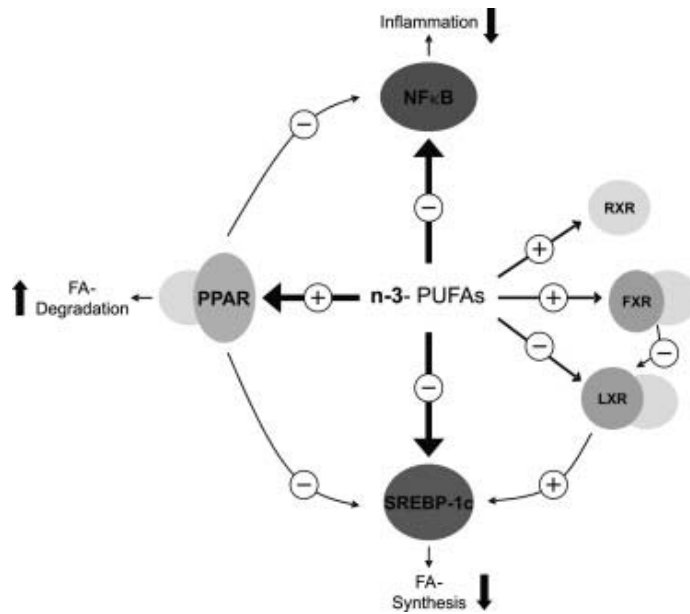


Figure 4. Nuclear receptors influenced by ω -3 FA. FXR, farnesoid X receptor; LXR, liver X receptor; NF κ B, nuclear factor k B; PPAR, peroxysome proliferators activated receptor; RXR, retinoid X receptor; SREBP-1c, sterol regulatory element binding protein 1c. Adapted from Schmitz et al. 2008.

The question remains how different types of FA control the expression of genes and a direct examination of the effect of each individual FA on porcine muscle FA composition is needed. However, well-controlled nutrigenomic studies on the regulation of gene transcription by FA are limited in livestock and even in humans (Afman et al., 2012). Integration among research groups with complementary knowledge and expertise such as molecular biologists, quantitative geneticists, nutritionists, microbiologists and bioinformatics are required to provide insights into the complexity of biological processes and therefore to establish more appropriate strategies in animal breeding programs.

CHAPTER IV

CONCLUSIONS

CHAPTER 4. CONCLUSIONS

- 1 - The first whole-genome map of CNVRs in pigs was generated using the 60 k SNP chip. Forty-nine CNVRs were identified as segregating in different pig populations. Five out of seven analyzed CNVRs were validated by real time quantitative PCR.
- 2 - A GWAS revealed a complex genetic architecture for IMF FA composition and indices of FA metabolism. A total of 43 chromosomal regions were associated with these traits in an Iberian \times Landrace cross. Twenty-three (53.5%) of these intervals contain QTLs for lipid traits reported in previous studies, suggesting a pleiotropic effect, mainly for QTLs in SSC4 and SSC8.
- 3 - An analysis of the liver transcriptome revealed a total of 55 genes differentially expressed between two groups of pigs with extreme phenotypes for muscle FA composition. Functional analysis revealed that these genes belong to biological functions, canonical pathways and three linked gene-networks related to lipid and FA metabolism.
- 4 - Using a holistic approach, a co-association gene-network containing 1,096 nodes linked by 111,198 interactions was inferred. Pathway and network analyses pointed to a trio of TF (*NCOA2*, *EP300*, *FHL2*) as key regulators of FA composition. Co-associations observed at the genetic level were partially validated through co-expression at the transcriptomic level in liver and adipose tissue. These results, also supported by bibliography, suggest a co-operative role of three TF in the transcriptional regulation of IMF FA composition and metabolism in pigs.

CHAPTER V

REFERENCES

CHAPTER 5. REFERENCES

- Abdolreza Mohammadi, and E. C. Wit. 2012. Gaussian graphical model determination based on birth-death mcmc inference. arXiv:1210.5371v4
- Afman, L. A., and M. Muller. 2012. Human nutrigenomics of gene regulation by dietary fatty acids. *Progress in Lipid Research* 51: 63-70.
- Alkan, C., J. M. Kidd, T. Marques-Bonet, G. Aksay, F. Antonacci, F. Hormozdiari, J. O. Kitzman, C. Baker, M. Malig, O. Mutlu, S. C. Sahinalp, R. A. Gibbs, and E. E. Eichler. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41: 1061-1067.
- Andersson, L., C. S. Haley, H. Ellegren, S. A. Knott, M. Johansson, K. Andersson, L. Andersson-Eklund, I. Edfors-Lilja, M. Fredholm, I. Hansson, and et al. 1994. Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263: 1771-1774.
- Aravena, A., D. Eveillard, A. Maass, and A. Siegel. 2013. A logics-based integrative approach to decipher putative regulatory relationships inferred from genomic and transcriptomic data Meeting on Gene Network Inference with Systems genetic data and beyond, Paris. 28-29th March.
- Archibald, A. L., L. Bolund, C. Churcher, M. Fredholm, M. A. Groenen, B. Harlizius, K. T. Lee, D. Milan, J. Rogers, M. F. Rothschild, H. Uenishi, J. Wang, and L. B. Schook. 2010. Pig genome sequence--analysis and publication strategy. *BMC Genomics* 11: 438.
- Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A. K. Wintero, and et al. 1995. The pigmap consortium linkage map of the pig (*sus scrofa*). *Mamm Genome* 6: 157-175.

Asselbergs, F., Y. Guo, E. P. A. van Iperen, S. Sivapalaratnam, V. Tragante, M. B. Lanktree, L. A. Lange, B. Almoguera, Y. E. Appelman, J. Barnard, J. Baumert, A. L. Beitelshes, T. R. Bhangale, Y.-D. I. Chen, T. R. Gaunt, Y. Gong, J. C. Hopewell, T. Johnson, M. E. Kleber, T. Y. Langae, M. Li, Y. R. Li, K. Liu, C. W. McDonough, M. F. L. Meijs, R. P. S. Middelberg, K. Musunuru, C. P. Nelson, J. R. O'Connell, S. Padmanabhan, J. S. Pankow, N. Pankratz, S. Rafelt, R. Rajagopalan, S. P. R. Romaine, N. J. Schork, J. Shaffer, H. Shen, E. N. Smith, S. E. Tischfield, P. J. van der Most, J. V. van Vliet-Ostapchouk, N. Verweij, K. A. Volcik, L. Zhang, K. R. Bailey, K. M. Bailey, F. Bauer, J. M. A. Boer, P. S. Braund, A. Burt, P. R. Burton, S. G. Buxbaum, W. Chen, R. M. Cooper-DeHoff, L. A. Cupples, J. S. deJong, C. Delles, D. Duggan, M. Fornage, C. E. Furlong, N. Glazer, J. G. Gums, C. Hastie, M. V. Holmes, T. Illig, S. A. Kirkland, M. Kivimaki, R. Klein, B. E. Klein, C. Kooperberg, K. Kottke-Marchant, M. Kumari, A. Z. LaCroix, L. Mallela, G. Murugesan, J. Ordovas, W. H. Ouwehand, W. S. Post, R. Saxena, H. Scharnagl, P. J. Schreiner, T. Shah, D. C. Shields, D. Shimbo, S. R. Srinivasan, R. P. Stolk, D. I. Swerdlow, H. A. Taylor Jr, E. J. Topol, E. Toskala, J. L. van Pelt, J. van Setten, S. Yusuf, J. C. Whittaker, A. H. Zwiderman, S. S. Anand, A. J. Balmforth, G. S. Berenson, C. R. Bezzina, B. O. Boehm, E. Boerwinkle, J. P. Casas, M. J. Caulfield, R. Clarke, J. M. Connell, K. J. Cruickshanks, K. W. Davidson, I. N. M. Day, P. I. W. de Bakker, P. A. Doevendans, A. F. Dominiczak, A. S. Hall, C. A. Hartman, C. Hengstenberg, H. L. Hillege, M. H. Hofker, S. E. Humphries, G. P. Jarvik, J. A. Johnson, B. M. Kaess, S. Kathiresan, W. Koenig, D. A. Lawlor, W. M. M. Melander, B. D. Mitchell, G. W. Montgomery, P. B. Munroe, S. S. Murray, S. J. Newhouse, N. C. Onland-Moret, N. Poulter, B. Psaty, S. Redline, S. S. Rich, J. I. Rotter, H. Schunkert, P. Sever, A. R. Shuldiner, R. L.

- Silverstein, A. Stanton, B. Thorand, M. D. Trip, M. Y. Tsai, P. van der Harst, E. van der Schoot, Y. T. van der Schouw, W. M. M. Verschuren, H. Watkins, A. A. M. Wilde, B. H. R. Wolffenbuttel, J. B. Whitfield, G. K. Hovingh, C. M. Ballantyne, C. Wijmenga, M. P. Reilly, N. G. Martin, J. G. Wilson, D. J. Rader, N. J. Samani, A. P. Reiner, R. A. Hegele, J. J. P. Kastelein, A. D. Hingorani, P. J. Talmud, H. Hakonarson, C. C. Elbers, B. J. Keating, and F. Drenos. 2012. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *The American Journal of Human Genetics* 91: 823-838.
- Badaoui, B., C. Tuggle, Z. Hu, J. Reecy, T. Ait-Ali, A. Anselmo, and S. Botti. 2013. Pig immune response to general stimulus and to porcine reproductive and respiratory syndrome virus infection: A meta-analysis approach. *BMC Genomics* 14: 220.
- Bansal, M., G. Della Gatta, and D. di Bernardo. 2006. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* 22: 815-822.
- Burgos-Paz, W., C. A. Souza, H. J. Megens, Y. Ramayo-Caldas, M. Melo, C. Lemus-Flores, E. Caal, H. W. Soto, R. Martinez, L. A. Alvarez, L. Aguirre, V. Iniguez, M. A. Revidatti, O. R. Martinez-Lopez, S. Llambi, A. Esteve-Codina, M. C. Rodriguez, R. P. Crooijmans, S. R. Paiva, L. B. Schook, M. A. Groenen, and M. Perez-Enciso. 2013. Porcine colonization of the americas: A 60k snp story. *Heredity (Edinb)* 110: 321-330.
- Califano, A., A. J. Butte, S. Friend, T. Ideker, and E. Schadt. 2012. Leveraging models of cell regulation and gwas data in integrative network-based association studies. *Nat Genet* 44: 841-847.

- Cameron, N. D. 1993. Selection for meat quality: Objectives and criteria. *Pig news and information* 14:161-168.
- Carrapiso, A. I., F. Bonilla, and C. García. 2003. Effect of crossbreeding and rearing system on sensory characteristics of iberian ham. *Meat Science* 65: 623-629.
- Carter, N. P. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 39: S16-21.
- Casellas, J., J. L. Noguera, J. Reixach, I. Diaz, M. Amills, and R. Quintanilla. 2010. Bayes factor analyses of heritability for serum and muscle lipid traits in duroc pigs. *Journal of Animal Science* 88: 2246-2254.
- Chen, C., H. Ai, J. Ren, W. Li, P. Li, R. Qiao, J. Ouyang, M. Yang, J. Ma, and L. Huang. 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.
- Chen, C., R. Qiao, R. Wei, Y. Guo, H. Ai, J. Ma, J. Ren, and L. Huang. 2012. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC Genomics* 13: 733.
- Chizzolini, R., E. Novelli, and E. Zanardi. 1998. Oxidation in traditional mediterranean meat products. *Meat Science* 49, Supplement 1: S87-S99.
- Chizzolini, R., E. Zanardi, V. Dorigoni, and S. Ghidini. 1999. Calorific value and cholesterol content of normal and low-fat meat and meat products. *Trends in Food Science & Technology* 10: 119-128.
- Cho, D. Y., Y. A. Kim, and T. M. Przytycka. 2012. Chapter 5: Network biology approach to complex diseases. *PLoS Comput Biol* 8: e1002820.
- Clayton-Smith, J., C. Giblin, R. A. Smith, C. Dunn, and L. Willatt. 2010. Familial 3q29 microdeletion syndrome providing further evidence of involvement of the 3q29

- region in bipolar disorder. *Clinical Dysmorphology* 19: 128-132
110.1097/MCD.1090b1013e32833a32831e32833c.
- Clop, A., C. Ovilo, M. Perez-Enciso, A. Cercos, A. Tomas, A. Fernandez, A. Coll, J. Folch, C. Barragan, I. Diaz, M. Oliver, L. Varona, L. Silio, A. Sanchez, and J. Noguera. 2003a. Detection of qtl affecting fatty acid composition in the pig. *Mammalian Genome* 14: 650-656.
- Clop, A., C. Ovilo, M. Perez-Enciso, A. Cercos, A. Tomas, A. Fernandez, A. Coll, J. M. Folch, C. Barragan, I. Diaz, M. A. Oliver, L. Varona, L. Silio, A. Sanchez, and J. L. Noguera. 2003b. Detection of qtl affecting fatty acid composition in the pig. *Mamm Genome* 14: 650-656.
- Cooper, G. M., T. Zerr, J. M. Kidd, E. E. Eichler, and D. A. Nickerson. 2008. Systematic assessment of copy number variant detection via genome-wide snp genotyping. *Nat Genet* 40: 1199-1203.
- Coppieters, W., A. Van de Weghe, A. Depicker, J. Coppieters, L. Peelman, A. Van Zeveren, and Y. Bouquet. 1995. Polymorphic cac/t repetitive sequences in the pig genome 1. *Anim Genet* 26: 327-330.
- Corominas, J., Y. Ramayo-Caldas, A. Puig-Oliveras, J. Estellé, A. Castelló, E. Alves, R. N. Pena, M. Ballester, and J. M. Folch. 2013a. Differential transcriptome analysis in adipose tissue of pigs phenotypically extremes for intramuscular fatty acid composition. Submitted.
- Corominas, J., Y. Ramayo-Caldas, A. Puig-Oliveras, D. Perez-Montarelo, J. L. Noguera, J. M. Folch, and M. Ballester. 2013b. Polymorphism in the *elovl6* gene is associated with a major qtl effect on fatty acid composition in pigs. *PLoS ONE* 8: e53687.

- Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12: 499-510.
- Davies, W., B. Hoyheim, B. Chaput, A. L. Archibald, and G. Frelat. 1994. Characterization of microsatellites from flow-sorted porcine chromosome 13. *Mamm Genome* 5: 707-711.
- De Smet, R., and K. Marchal. 2010. Advantages and limitations of current network inference methods. *Nat Rev Microbiol* 8: 717-729.
- Dekkers, J. C. M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science* 82: E313-E328.
- Dohm, J. C., C. Lottaz, T. Borodina, and H. Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* 36: e105.
- Druet, T., and M. Georges. 2009. A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 184: 789-798.
- Duijvesteijn, N., E. Knol, J. Merks, R. Crooijmans, M. Groenen, H. Bovenhuis, and B. Harlizius. 2010. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genetics* 11: 42.
- Dupuis, M. C., Z. Zhang, T. Druet, J. M. Denoix, C. Charlier, P. Lekeux, and M. Georges. 2011. Results of a haplotype-based gwas for recurrent laryngeal neuropathy in the horse. *Mamm Genome* 22: 613-620.
- Echard, G., D. Milan, M. Yerle, Y. Lahbib-Mansais, and J. Gellin. 1992. The gene map of the pig (*sus scrofa domestica* l.): A review. *Cytogenet Cell Genet* 61: 146-151.

- Emilsson, V., G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, and K. Stefansson. 2008. Genetics of gene expression and its effect on disease. *Nature* 452: 423-428.
- Estelle, J., A. I. Fernandez, M. Perez-Enciso, A. Fernandez, C. Rodriguez, A. Sanchez, J. L. Noguera, and J. M. Folch. 2009. A non-synonymous mutation in a conserved site of the *mtp* gene is strongly associated with protein activity and fatty acid profile in pigs. *Anim Genet* 40: 813-820.
- Esteve-Codina, A., R. Kofler, N. Palmieri, G. Bussotti, C. Notredame, and M. Perez-Enciso. 2011. Exploring the gonad transcriptome of two extreme male pigs with *rna-seq*. *BMC Genomics* 12: 552.
- Estévez, M., D. Morcuende, and R. Cava López. 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 Science* 64: 499-506.
- FAO. 2010. Fats and fatty acids in human nutrition. Report of expert consultation
- Fernandez, A. I., D. Perez-Montarelo, C. Barragan, Y. Ramayo-Caldas, N. Ibanez-Escriche, A. Castello, J. L. Noguera, L. Silio, J. M. Folch, and M. C. Rodriguez. 2012. Genome-wide linkage analysis of qtl for growth and body composition employing the *porcinesnp60* beadchip. *BMC Genet* 13: 41.
- Fernandez-Rodriguez, A., M. Munoz, A. Fernandez, R. N. Pena, A. Tomas, J. L. Noguera, C. Ovilo, and A. I. Fernandez. 2011. Differential gene expression in

- ovaries of pregnant pigs with high and low prolificacy levels and identification of candidate genes for litter size. *Biology of Reproduction* 84: 299-307.
- Ferraz, A., A. Ojeda, M. Lopez-Bejar, L. Fernandes, A. Castello, J. Folch, and M. Perez-Enciso. 2008. Transcriptome architecture across tissues in the pig. *BMC Genomics* 9: 173.
- Filkov, V. 2005. Identifying gene regulatory networks from gene expression data. In: S. Aluru (ed.) *Handbook of computational molecular biology*. p 1104.
- Flicek, P., M. R. Amode, D. Barrell, K. Beal, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. Kahari, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, P. Larsson, I. Longden, W. McLaren, B. Overduin, B. Pritchard, H. S. Riat, D. Rios, G. R. Ritchie, M. Ruffier, M. Schuster, D. Sobral, G. Spudich, Y. A. Tang, S. Trevanion, J. Vandrovcova, A. J. Vilella, S. White, S. P. Wilder, A. Zadissa, J. Zamora, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J. Hubbard, A. Parker, G. Proctor, J. Vogel, and S. M. Searle. 2011. Ensembl 2011. *Nucleic Acids Res* 39: D800-806.
- Fontanesi, L., G. Galimberti, D. G. Caló, M. Colombo, A. Astolfi, S. Formica, and V. Russo. 2011. Microarray gene expression analysis of porcine skeletal muscle sampled at several post mortem time points. *Meat Science* 88: 604-609.
- Fontanesi, L., G. Schiavo, G. Galimberti, D. G. Calo, E. Scotti, P. L. Martelli, L. Buttazzoni, R. Casadio, and V. Russo. 2012. A genome wide association study for backfat thickness in italian large white pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* 13: 583.
- Fortes, M. R., A. Reverter, Y. Zhang, E. Collis, S. H. Nagaraj, N. N. Jonsson, K. C. Prayaga, W. Barris, and R. J. Hawken. 2010. Association weight matrix for the

- genetic dissection of puberty in beef cattle. *Proc Natl Acad Sci U S A* 107: 13642-13647.
- Fortes, M. R. S., A. Reverter, S. H. Nagaraj, Y. Zhang, N. N. Jonsson, W. Barris, S. Lehnert, G. B. Boe-Hansen, and R. J. Hawken. 2011. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. *Journal of Animal Science* 89: 1669-1683.
- Fortes, M. R. S., W. M. Snelling, A. Reverter, S. H. Nagaraj, S. A. Lehnert, R. J. Hawken, K. L. DeAtley, S. O. Peters, G. A. Silver, G. Rincon, J. F. Medrano, A. Islas-Trejo, and M. G. Thomas. 2012. Gene network analyses of first service conception in brangus heifers: Use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *Journal of Animal Science* 90: 2894-2906.
- Foulkes, A. S. 2009. *Applied statistical genetics with r. For population-based association studies*, New York.
- Freeman, T. C., A. Ivens, J. K. Baillie, D. Beraldi, M. W. Barnett, D. Dorward, A. Downing, L. Fairbairn, R. Kapetanovic, S. Raza, A. Tomoiu, R. Alberio, C. Wu, A. I. Su, K. M. Summers, C. K. Tuggle, A. L. Archibald, and D. A. Hume. 2012. A gene expression atlas of the domestic pig. *BMC Biol* 10: 90.
- Garber, M., M. G. Grabherr, M. Guttman, and C. Trapnell. 2011. Computational methods for transcriptome annotation and quantification using rna-seq. *Nat Meth* 8: 469-477.
- Gaspa, D. B., C. S. Novell, A. C. Fabregat, J. Á. Rodríguez, E. G. Hernández, K. L. Romero, L. N. Solans, D. C. Tinoco, G. B. Abill, F. R. Fosas, and A. C. Ribalta. 2011. Informe del sector porcino. In: A. i. M. N. Departament d'Agricultura Ramaderia Pesca (ed.). Mayo.

- Georges, M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annu Rev Genomics Hum Genet* 8: 131-162.
- Glancy, M., A. Barnicoat, R. Vijeratnam, S. de Souza, J. Gilmore, S. Huang, V. K. Maloney, N. S. Thomas, D. J. Bunyan, A. Jackson, and J. C. Barber. 2009. Transmitted duplication of 8p23.1-8p23.2 associated with speech delay, autism and learning difficulties. *Eur J Hum Genet* 17: 37-43.
- Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 10: 381-391.
- Gregersen, V., L. Conley, K. Sorensen, B. Guldbrandtsen, I. Velander, and C. Bendixen. 2012. Genome-wide association scan and phased haplotype construction for quantitative trait loci affecting boar taint in three pig breeds. *BMC Genomics* 13: 22.
- Grindflek, E., S. Lien, H. Hamland, M. Hansen, M. Kent, M. van Son, and T. Meuwissen. 2011. Large scale genome-wide association and ldl mapping study identifies qtls for boar taint and related sex steroids. *BMC Genomics* 12: 362.
- Groenen, M. A., D. Ruyter, E. J. Verstege, M. de Vries, and J. J. van der Poel. 1995. Development and mapping of ten porcine microsatellite markers. *Anim Genet* 26: 115-118.
- Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H.-J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. F. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S.-H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. M. A. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z.-Q. Du, K. Eversole, J. Fadista, S. Fairley,

T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. R. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z.-L. Hu, S. J. Humphray, T. Hunt, H. Hornshoj, J.-T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J.-H. Kim, K.-W. Kim, T.-H. Kim, G. Larson, K. Lee, K.-T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. Truong Nguyen, P. Ni, S.-J. Oh, S. Onteru, F. Panitz, E.-W. Park, H.-S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, and L. B. Schook. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393-398.

Gu, Q., S. Nagaraj, N. Hudson, B. Dalrymple, and A. Reverter. 2011. Genome-wide patterns of promoter sharing and co-expression in bovine skeletal muscle. *BMC Genomics* 12: 23.

Guillevic, M., M. Kouba, and J. Mouro. 2009. Effect of a linseed diet or a sunflower diet on performances, fatty acid composition, lipogenic enzyme activities and stearoyl-coa-desaturase activity in the pig. *Livestock Science* 124: 288-294.

Guryev, V., K. Saar, T. Adamovic, M. Verheul, S. A. A. C. van Heesch, S. Cook, M. Pravenec, T. Aitman, H. Jacob, J. D. Shull, N. Hubner, and E. Cuppen. 2008. Distribution and functional impact of DNA copy number variation in the rat. *Nat Genet* 40: 538-545.

- Haley, C. S., A. L. Archibald, L. Andersson, A. A. Bosma, W. Davies, M. Fredholm, H. Geldermann, M. Groenen, I. Gustavsson, L. Ollivier, E. M. Tucker, and A. Van de Weghe. 1990. The pig gene mapping project – pigmap. In: 4th World Congress on Genetics Applied to Livestock Production, Edinburgh. p 67–70.
- Harhay, G. P., T. P. Smith, L. J. Alexander, C. D. Haudenschild, J. W. Keele, L. K. Matukumalli, S. G. Schroeder, C. P. Van Tassell, C. R. Gresham, S. M. Bridges, S. C. Burgess, and T. S. Sonstegard. 2010. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. *Genome Biol* 11: R102.
- Hawken, R. J., Y. D. Zhang, M. R. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A. Reverter, and S. A. Lehnert. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. *J Anim Sci* 90: 1398-1410.
- Hayes, B. 2007. Qtl mapping, mas, and genomic selection p118, Iowa State University.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell, T. S. Sonstegard, and M. E. Goddard. 2009. A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS ONE* 4: e6676.
- Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13: 635-643.
- Henrichsen, C. N., E. Chaignat, and A. Reymond. 2009a. Copy number variants, diseases and gene expression. *Hum Mol Genet* 18: R1-8.

- Henrichsen, C. N., N. Vinckenbosch, S. Zollner, E. Chagnat, S. Pradervand, F. Schutz, M. Ruedi, H. Kaessmann, and A. Reymond. 2009b. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* 41: 424-429.
- Hu, Z. L., C. A. Park, X. L. Wu, and J. M. Reecy. 2013. Animal qtladb: An improved database tool for livestock animal qtl/association data dissemination in the post-genome era. *Nucleic Acids Res* 41: D871-879.
- Hu, Z. L., and J. M. Reecy. 2007. Animal qtladb: Beyond a repository. A public platform for qtl comparisons and integration with diverse types of structural genomic information. *Mamm Genome* 18: 1-4.
- Huang, W., B. W. Kirkpatrick, G. J. M. Rosa, and H. Khatib. 2010. A genome-wide association study using selective DNA pooling identifies candidate markers for fertility in holstein cattle. *Animal Genetics* 41: 570-578.
- Hudson, N., B. Dalrymple, and A. Reverter. 2012. Beyond differential expression: The quest for causal mutations and effector molecules. *BMC Genomics* 13: 356.
- Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in chinese holstein population. *PLoS ONE* 5: e13661.
- Jiang, L., P. Sorensen, B. Thomsen, S. M. Edwards, A. Skarman, C. M. Rontved, M. S. Lund, and C. T. Workman. 2011. Gene prioritization for livestock diseases by data integration. *Physiological Genomics* 44: 305-317.
- Jung, W. Y., S. G. Kwon, M. Son, E. S. Cho, Y. Lee, J. H. Kim, B. W. Kim, H. Park da, J. H. Hwang, T. W. Kim, H. C. Park, B. Y. Park, J. S. Choi, K. K. Cho, K. H. Chung, Y. M. Song, I. S. Kim, S. K. Jin, D. H. Kim, S. W. Lee, K. W. Lee, W. Y. Bang, and C. W. Kim. 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.

- Karim, L., H. Takeda, L. Lin, T. Druet, J. A. Arias, D. Baurain, N. Cambisano, S. R. Davis, F. Farnir, B. Grisart, B. L. Harris, M. D. Keehan, M. D. Littlejohn, R. J. Spelman, M. Georges, and W. Coppieters. 2011. Variants modulating the expression of a chromosome domain encompassing *plag1* influence bovine stature. *Nat Genet* 43: 405-413.
- Katan, M., P. Zock, and R. Mensink. 1994. Effects of fats and fatty acids on blood lipids in humans: An overview. *The American Journal of Clinical Nutrition* 60: 1017S-1022S.
- Klein, C., K. Lohmann, and A. Ziegler. 2012. The promise and limitations of genome-wide association studies. *JAMA*: 1-2.
- Klein, R. J., C. Zeiss, E. Y. Chew, J.-Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh. 2005. Complement factor h polymorphism in age-related macular degeneration. *Science* 308: 385-389.
- Kogelman, L., K. Byrne, T. Vuocolo, N. Watson-Haigh, H. Kadarmideen, J. Kijas, H. Oddy, G. Gardner, C. Gondro, and R. Tellam. 2011. Genetic architecture of gene expression in ovine skeletal muscle. *BMC Genomics* 12: 607.
- Kuehn, C., C. Edel, R. Weikard, and G. Thaller. 2007. Dominance and parent-of-origin effects of coding and non-coding alleles at the acylcoa-diacylglycerol-acyltransferase (*dgat1*) gene on milk production traits in german holstein cows. *BMC Genetics* 8: 62.
- Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics* 121: 185-199.
- Langfelder, P., and S. Horvath. 2008. Wgcna: An r package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559.

- Larson, G., K. Dobney, U. Albarella, M. Fang, E. Matisoo-Smith, J. Robins, S. Lowden, H. Finlayson, T. Brand, E. Willerslev, P. Rowley-Conwy, L. Andersson, and A. Cooper. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307: 1618-1621.
- Ledur, M. C., N. Navarro, and M. Perez-Enciso. 2010. Large-scale snp genotyping in crosses between outbred lines: How useful is it? *Heredity* 105: 173-182.
- Li, Y., S. Mei, X. Zhang, X. Peng, G. Liu, H. Tao, H. Wu, S. Jiang, Y. Xiong, and F. Li. 2012. Identification of genome-wide copy number variations among diverse pig breeds by array cgh. *BMC Genomics* 13: 725.
- Lim, D., N. Kim, H. Park, S. Lee, Y. Cho, S. Oh, T. Kim, and H. Kim. 2011. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *Int J Biol Sci* 7(7):992-1002. doi:10.7150/ijbs.7.992.
- Luo, W., S. Chen, D. Cheng, L. Wang, Y. Li, X. Ma, X. Song, X. Liu, W. Li, J. Liang, H. Yan, K. Zhao, C. Wang, L. Wang, and L. Zhang. 2012a. Genome-wide association study of porcine hematological parameters in a large white × minzhu f2 resource population. *Int J Biol Sci* 8(6): 11.
- Luo, W., D. Cheng, S. Chen, L. Wang, Y. Li, X. Ma, X. Song, X. Liu, W. Li, J. Liang, H. Yan, K. Zhao, C. Wang, and L. Zhang. 2012b. Genome-wide association analysis of meat quality traits in a porcine large white x minzhu intercross population. *Int J Biol Sci* 8: 580-595.
- Mackay, T. F., E. A. Stone, and J. F. Ayroles. 2009. The genetics of quantitative traits: Challenges and prospects. *Nat Rev Genet* 10: 565-577.
- Maher, B. 2012. Encode: The human encyclopaedia. *Nature* 489: 46-48.
- Mai, M. D., G. Sahana, F. B. Christiansen, and B. Guldbbrandtsen. 2009. A genome-wide association study for milk production traits in danish jersey cattle using a

- 50k single nucleotide polymorphism chip. *Journal of Animal Science* 88: 3522-3528.
- Manunza, A., A. Zidi, S. Yeghoyan, V. A. Balteanu, T. C. Carsai, O. Scherbakov, O. Ramirez, S. Eghbalsaied, A. Castello, A. Mercade, and M. Amills. 2013. A high throughput genotyping approach reveals distinctive autosomal genetic signatures for european and near eastern wild boar. *PLoS ONE* 8: e55891.
- Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad. 2008. Rna-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18: 1509-1517.
- Marklund, S., J. Kijas, H. Rodriguez-Martinez, L. Ronnstrand, K. Funa, M. Moller, D. Lange, I. Edfors-Lilja, and L. Andersson. 1998. Molecular basis for the dominant white phenotype in the domestic pig. *Genome Research* 8: 826-833.
- Melo, C. T. 2012. Caracterización de genes candidatos para caracteres lipídicos e identificación de copy number variation en una población porcina duroc, Universitat Autònoma de Barcelona, Barcelona.
- Meuwissen, T., and M. Goddard. 2004. Mapping multiple qtl using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics Selection Evolution* 36: 261 - 279.
- Meuwissen, T. H., and M. E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155: 421-430.
- Moller, M., R. Chaudhary, E. Hellmén, B. Höyheim, B. Chowdhary, and L. Andersson. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the kit gene encoding the mast/stem cell growth factor receptor. *Mammalian Genome* 7: 822-830.

- Morais, S., J. Taggart, D. Guy, J. Bell, and D. Tocher. 2012. Hepatic transcriptome analysis of inter-family variability in flesh n-3 long-chain polyunsaturated fatty acid content in atlantic salmon. *BMC Genomics* 13: 410.
- Mordelet, F., and J. P. Vert. 2013. Supervised inference of gene regulatory networks from positive and unlabeled examples. *Methods Mol Biol* 939: 47-58.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods* 5: 621-628.
- Nafikov, R., and D. Beitz. 2007. Carbohydrate and lipid metabolism in farm animals. *J Nutr. Mar*;137(3):: 702-705.
- Nordborg, M., and S. Tavaré. 2002. Linkage disequilibrium: What history has to tell us. *Trends in Genetics* 18: 83-90.
- Ntawubizi, M., E. Colman, S. Janssens, K. Raes, N. Buys, and S. De Smet. 2010. Genetic parameters for intramuscular fatty acid composition and metabolism in pigs. *J Anim Sci* 88: 1286-1294.
- O'Hea, E., and G. Leveille. 1969. Significance of adipose tissue and liver as sites of fatty acid synthesis in the pig and the efficiency of utilization of various substrates for lipogenesis. *J Nutr.* Nov 99(3): 338-344.
- Okada, T., N. Furuhashi, Y. Kuromori, M. Miyashita, F. Iwata, and K. Harada. 2005. Plasma palmitoleic acid content and obesity in children. *The American Journal of Clinical Nutrition* 82: 747-750.
- Olsson, M., J. R. S. Meadows, K. Truve, G. Rosengren Pielberg, F. Puppò, E. Mauceli, J. Quilez, N. Tonomura, G. Zanna, M. J. Docampo, A. Bassols, A. C. Avery, E. K. Karlsson, A. Thomas, D. L. Kastner, E. Bongcam-Rudloff, M. T. Webster, A. Sanchez, Ö. Hedhammar, E. F. Remmers, L. Andersson, L. Ferrer, L. Tintle, and K. Lindblad-Toh. 2011. A novel unstable duplication upstream of *has2*

- predisposes to a breed-defining skin phenotype and a periodic fever syndrome in chinese shar-pei dogs. *PLoS Genet* 7: e1001332.
- Onteru, S. K., B. Fan, Z. Q. Du, D. J. Garrick, K. J. Stalder, and M. F. Rothschild. 2011. A whole-genome association study for pig reproductive traits. *Anim Genet* 43: 18-26.
- Palsson, B. O. 2006. *Systems biology. Properties of reconstructed networks.* April 2006, New York.
- Pascual, J., M. Rafecas, M. Canela, J. Boatella, R. Bou, A. Barroeta, and R. Codony. 2007. Effect of increasing amounts of a linoleic-rich dietary fat on the fat composition of four pig breeds. Part ii: Fatty acid composition in muscle and fat tissues. *Food Chemistry* 100: 1639-1648.
- Perez-Enciso, M., A. Clop, J. L. Noguera, C. Ovilo, A. Coll, J. M. Folch, D. Babot, J. Estany, M. A. Oliver, I. Diaz, and A. Sanchez. 2000. A qtl on pig chromosome 4 affects fatty acid metabolism: Evidence from an iberian by landrace intercross. *J Anim Sci* 78: 2525-2531.
- Perez-Enciso, M., A. L. Ferraz, A. Ojeda, and M. Lopez-Bejar. 2009. Impact of breed and sex on porcine endocrine transcriptome: A bayesian biometrical analysis. *BMC Genomics* 10: 89.
- Perez-Enciso, M., and L. Ferretti. 2010. Massive parallel sequencing in animal genetics: Wherefroms and wheretos. *Anim Genet* 41: 561-569.
- Perez-Montarelo, D., N. J. Hudson, A. I. Fernandez, Y. Ramayo-Caldas, B. P. Dalrymple, and A. Reverter. 2012. Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. *PLoS ONE* 7: e46159.
- Pielberg, G., C. Olsson, A. Syvanen, and L. Andersson. 2002. Unexpectedly high allelic diversity at the kit locus causing dominant white color in the domestic pig. *Genetics* 160: 305 - 311.

- Ponsuksili, S., E. Murani, C. Phatsara, E. Jonas, C. Walz, M. Schwerin, K. Schellander, and K. Wimmers. 2008. Expression profiling of muscle reveals transcripts differentially expressed in muscle that affect water-holding capacity of pork. *J Agric Food Chem* 56: 10311-10317.
- Poudyal, H., S. K. Panchal, V. Diwan, and L. Brown. 2011. Omega-3 fatty acids and metabolic syndrome: Effects and emerging mechanisms of action. *Progress in Lipid Research* 50: 372-387.
- Puig-Oliveras, A., Y. Ramayo-Caldas, J. Corominas, J. Estellé, A. I. Fernandez, J. Casellas, M. Ballester, and J. M. Folch. 2013. Differences in swine muscle transcriptome among animals phenotypically extremes for fatty acid composition. Submitted.
- Ramayo-Caldas, Y., A. Castello, R. Pena, E. Alves, A. Mercade, C. Souza, A. Fernandez, M. Perez-Enciso, and J. Folch. 2010. Copy number variation in the porcine genome inferred from a 60 k snp beadchip. *BMC Genomics* 11: 593.
- Ramayo-Caldas, Y., N. Mach, A. Esteve-Codina, J. Corominas, A. Castello, M. Ballester, J. Estelle, N. Ibanez-Escriche, A. I. Fernandez, M. Perez-Enciso, and J. M. Folch. 2012a. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics* 13: 547.
- Ramayo-Caldas, Y., A. Mercade, A. Castello, B. Yang, C. Rodriguez, E. Alves, I. Diaz, N. Ibanez-Escriche, J. L. Noguera, M. Perez-Enciso, A. I. Fernandez, and J. M. Folch. 2012b. Genome-wide association study for intramuscular fatty acid composition in an iberian x landrace cross. *J Anim Sci* 90: 2883-2893.
- Ravnan, J. B., J. H. Tepperberg, P. Papenhausen, A. N. Lamb, J. Hedrick, D. Eash, D. H. Ledbetter, and C. L. Martin. 2006. Subtelomere fish analysis of 11 688 cases: An evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J Med Genet* 43: 478-489.

- Ren, J., H. Mao, Z. Zhang, S. Xiao, N. Ding, and L. Huang. 2011. A 6-bp deletion in the *tyrp1* gene causes the brown colouration phenotype in chinese indigenous pigs. *Heredity (Edinb)* 106: 862-868.
- Reverter, A., and E. K. Chan. 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24: 2491-2497.
- Rosengren Pielberg, G., A. Golovko, E. Sundstrom, I. Curik, J. Lennartsson, M. H. Seltenhammer, T. Druml, M. Binns, C. Fitzsimmons, G. Lindgren, K. Sandberg, R. Baumung, M. Vetterlein, S. Stromberg, M. Grabherr, C. Wade, K. Lindblad-Toh, F. Ponten, C.-H. Heldin, J. Solkner, and L. Andersson. 2008. A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* 40: 1004-1009.
- Rothschild, M. F., and A. Ruvinsky. 2011. *The genetics of the pig*, 2nd edition.
- Rubin, C.-J., H.-J. Megens, A. M. Barrio, K. Maqbool, S. Sayyab, D. Schwochow, C. Wang, ñ. Carlborg, P. Jern, C. B. Jorgensen, A. L. Archibald, M. Fredholm, M. A. M. Groenen, and L. Andersson. 2012. Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences* 109: 19529-19536.
- Rudel, L. L., J. S. Parks, and J. K. Sawyer. 1995. Compared with dietary monounsaturated and saturated fat, polyunsaturated fat protects african green monkeys from coronary artery atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 15: 2101-2110.
- Sahana, G., B. Guldbrandsen, C. Bendixen, and M. S. Lund. 2010. Genome-wide association mapping for female fertility traits in danish and swedish holstein cattle. *Anim Genet* 41: 579-588.

- Sahana, G., V. Kadlecova, H. Hornshoj, B. Nielsen, and O. F. Christensen. 2013. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *J Anim Sci* 91: 1041-1050.
- Salmon Hillbertz, N. H. C., M. Isaksson, E. K. Karlsson, E. Hellmen, G. R. Pielberg, P. Savolainen, C. M. Wade, H. von Euler, U. Gustafson, A. Hedhammar, M. Nilsson, K. Lindblad-Toh, L. Andersson, and G. Andersson. 2007. Duplication of *fgf3*, *fgf4*, *fgf19* and *oraov1* causes hair ridge and predisposition to dermoid sinus in ridgeback dogs. *Nat Genet* 39: 1318-1320.
- Sartelet, A., T. Druet, C. Michaux, C. Fasquelle, S. Geron, N. Tamma, Z. Zhang, W. Coppieters, M. Georges, and C. Charlier. 2012. A splice site variant in the bovine *rnfl1* gene compromises growth and regulation of the inflammatory response. *PLoS Genet* 8: e1002581.
- Saxena, R., C. Elbers, Y. Guo, I. Peter, T. R. Gaunt, J. L. Mega, M. B. Lanktree, A. Tare, B. A. Castillo, Y. R. Li, T. Johnson, M. Bruinenberg, D. Gilbert-Diamond, R. Rajagopalan, B. F. Voight, A. Balasubramanyam, J. Barnard, F. Bauer, J. Baumert, T. Bhangale, B. O. Boehm, P. S. Braund, P. R. Burton, H. R. Chandrupatla, R. Clarke, R. M. Cooper-DeHoff, E. D. Crook, G. Davey-Smith, I. N. Day, A. de Boer, M. C. H. de Groot, F. Drenos, J. Ferguson, C. S. Fox, C. E. Furlong, Q. Gibson, C. Gieger, L. A. Gilhuijs-Pederson, J. T. Glessner, A. Goel, Y. Gong, S. F. A. Grant, D. E. Grobbee, C. Hastie, S. E. Humphries, C. E. Kim, M. Kivimaki, M. Kleber, C. Meisinger, M. Kumari, T. Y. Langae, D. A. Lawlor, M. Li, M. T. Lobbmeyer, A.-H. Maitland-van Zee, M. F. L. Meijis, C. M. Molony, D. A. Morrow, G. Murugesan, S. K. Musani, C. P. Nelson, S. J. Newhouse, J. R. O'Connell, S. Padmanabhan, J. Palmen, S. R. Patel, C. J. Pepine, M. Pettinger, T. S. Price, S. Rafelt, J. Ranchalis, A. Rasheed, E. Rosenthal, I. Ruczinski, S. Shah, H. Shen, G. n. Silbernagel, E. N. Smith, A.

W. M. Spijkerman, A. Stanton, M. W. Steffes, B. Thorand, M. Trip, P. van der Harst, D. L. van der A, E. P. A. van Iperen, J. van Setten, J. V. van Vliet-Ostaptchouk, N. Verweij, B. H. R. Wolffenbuttel, T. Young, M. H. Zafarmand, J. M. Zmuda, M. Boehnke, D. Altshuler, M. McCarthy, W. H. L. Kao, J. S. Pankow, T. P. Cappola, P. Sever, N. Poulter, M. Caulfield, A. Dominiczak, D. C. Shields, D. L. Bhatt, L. Zhang, S. P. Curtis, J. Danesh, J. P. Casas, Y. T. van der Schouw, N. C. Onland-Moret, P. A. Doevendans, G. W. Dorn II, M. Farrall, G. A. FitzGerald, A. Hamsten, R. Hegele, A. D. Hingorani, M. H. Hofker, G. S. Huggins, T. Illig, G. P. Jarvik, J. A. Johnson, O. H. Klungel, W. C. Knowler, W. Koenig, W. März, J. B. Meigs, O. Melander, P. B. Munroe, B. D. Mitchell, S. J. Bielinski, D. J. Rader, M. P. Reilly, S. S. Rich, J. I. Rotter, D. Saleheen, N. J. Samani, E. E. Schadt, A. R. Shuldiner, R. Silverstein, K. Kottke-Marchant, P. J. Talmud, H. Watkins, F. W. Asselbergs, P. I. W. de Bakker, J. McCaffery, C. Wijmenga, M. S. Sabatine, J. G. Wilson, A. Reiner, D. W. Bowden, H. Hakonarson, D. S. Siscovick, and B. J. Keating. 2012. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *The American Journal of Human Genetics* 90: 410-425.

Schadt, E. E., J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S. K. Sieberts, S. Monks, M. Reitman, C. Zhang, P. Y. Lum, A. Leonardson, R. Thieringer, J. M. Metzger, L. Yang, J. Castle, H. Zhu, S. F. Kash, T. A. Drake, A. Sachs, and A. J. Lusis. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 37: 710-717.

Schaschl, H., T. J. Aitman, and T. J. Vyse. 2009. Copy number variation in the human genome and its implication in autoimmunity. *Clin Exp Immunol* 156: 12-16.

- Schlechtriem, C., J. E. Bron, and D. R. Tocher. 2007. Inter-individual variation in total fatty acid compositions of flesh of atlantic salmon smolts-fed diets containing fish oil or vegetable oil. *Aquaculture Research* 38: 1045-1055.
- Schmitz, G., and J. Ecker. 2008. The opposing effects of n-3 and n-6 fatty acids. *Progress in Lipid Research* 47: 147-155.
- Schneider, J. F., L. A. Rempel, W. M. Snelling, R. T. Wiedmann, D. J. Nonneman, and G. A. Rohrer. 2012. Genome-wide association study of swine farrowing traits. Part ii: Bayesian analysis of marker data. *J Anim Sci* 90: 3360-3367.
- Sellier, P. 1998. Genetic improvement of the pig. In: C. International (ed.) *The genetics of the pig*. p 511-540., Wallingford, Oxon.
- Sellier, P., L. Maignel, and J. P. Bidanel. 2010. Genetic parameters for tissue and fatty acid composition of backfat, perirenal fat and longissimus muscle in large white and landrace pigs. *Animal* 4: 497-504.
- Serra, X., F. Gil, M. Pérez-Enciso, M. A. Oliver, J. M. Vázquez, M. Gispert, I. Díaz, F. Moreno, R. Latorre, and J. L. Noguera. 1998. A comparison of carcass, meat quality and histochemical characteristics of iberian (guadyerbas line) and landrace pigs. *Livestock Production Science* 56: 215-223.
- Sha, B.-Y., T.-L. Yang, L.-J. Zhao, X.-D. Chen, Y. Guo, Y. Chen, F. Pan, Z.-X. Zhang, S.-S. Dong, X.-H. Xu, and H.-W. Deng. 2009. Genome-wide association study suggested copy number variation may be associated with body mass index in the chinese population. *J Hum Genet* 54: 199-202.
- Sharp, A. J., H. C. Mefford, K. Li, C. Baker, C. Skinner, R. E. Stevenson, R. J. Schroer, F. Novara, M. De Gregori, R. Ciccone, A. Broomer, I. Casuga, Y. Wang, C. Xiao, C. Barbacioru, G. Gimelli, B. D. Bernardina, C. Torniero, R. Giorda, R. Regan, V. Murday, S. Mansour, M. Fichera, L. Castiglia, P. Failla, M. Ventura, Z. Jiang, G. M. Cooper, S. J. Knight, C. Romano, O. Zuffardi, C. Chen, C. E.

- Schwartz, and E. E. Eichler. 2008. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 40: 322-328.
- Sioutis, S., A. M. Coates, J. D. Buckley, T. W. Murphy, H. A. Channon, and P. R. C. Howe. 2008. N-3 enrichment of pork with fishmeal: Effects on production and consumer acceptability. *European Journal of Lipid Science and Technology* 110: 701-706.
- Snel, B., G. Lehmann, P. Bork, and M. A. Huynen. 2000. String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research* 28: 3442-3444.
- Stables, M. J., and D. W. Gilroy. 2011. Old and new generation lipid mediators in acute inflammation and resolution. *Progress in Lipid Research* 50: 35-51.
- Stanley, D., N. Watson-Haigh, C. J. Cowled, and R. Moore. 2013. Genetic architecture of gene expression in the chicken. *BMC Genomics* 14: 13.
- Stranger, B. E., E. A. Stahl, and T. Raj. 2011. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187: 367-383.
- Sun, W., N. Hudson, A. Reverter, A. Waardenberg, R. Tellam, T. Vuocolo, K. Byrne, and B. Dalrymple. 2012. An always correlated gene expression landscape for ovine skeletal muscle, lessons learnt from comparison with an "Equivalent" Bovine landscape. *BMC Research Notes* 5: 632.
- Sun, Y. V. 2012. Integration of biological networks and pathways with genetic association studies. *Hum Genet* 131: 1677-1686.
- Sutter, N. B., and E. A. Ostrander. 2004. Dog star rising: The canine genetic system. *Nat Rev Genet* 5: 900-910.
- Suzuki, K., M. Ishida, H. Kadowaki, T. Shibata, H. Uchida, and A. Nishida. 2006. Genetic correlations among fatty acid compositions in different sites of fat

- tissues, meat production, and meat quality traits in duroc pigs. *Journal of Animal Science* 84: 2026-2034.
- Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* 2: 125-141.
- Szklarczyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguetz, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. v. Mering. 2011. The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* 39: D561-D568.
- Tam, G. W., R. Redon, N. P. Carter, and S. G. Grant. 2009. The role of DNA copy number variation in schizophrenia. *Biol Psychiatry* 66: 1005-1012.
- Teo, S. M., Y. Pawitan, C. S. Ku, K. S. Chia, and A. Salim. 2012. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 28: 2711-2718.
- Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, C. T. Johansen, S. W. Fouchier, A. Isaacs, G. M. Peloso, M. Barbalic, S. L. Ricketts, J. C. Bis, Y. S. Aulchenko, G. Thorleifsson, M. F. Feitosa, J. Chambers, M. Orholm, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho, M. Jin Go, Y. Jin Kim, J.-Y. Lee, T. Park, K. Kim, X. Sim, R. Tzee-Hee Ong, D. C. Croteau-Chonka, L. A. Lange, J. D. Smith, K. Song, J. Hua Zhao, X. Yuan, J. a. Luan, C. Lamina, A. Ziegler, W. Zhang, R. Y. L. Zee, A. F. Wright, J. C. M. Witteman, J. F. Wilson, G. Willemsen, H. E. Wichmann, J. B. Whitfield, D. M. Waterworth, N. J. Wareham, G. Waeber, P. Vollenweider, B. F. Voight, V. Vitart, A. G. Uitterlinden, M. Uda, J. Tuomilehto, J. R. Thompson, T. Tanaka, I. Surakka, H. M. Stringham, T. D. Spector, N. Soranzo, J. H. Smit, J. Sinisalo, K. Silander, E. J. G. Sijbrands, A. Scuteri, J. Scott, D. Schlessinger, S. Sanna, V.

Salomaa, J. Saharinen, C. Sabatti, A. Ruukonen, I. Rudan, L. M. Rose, R. Roberts, M. Rieder, B. M. Psaty, P. P. Pramstaller, I. Pichler, M. Perola, B. W. J. H. Penninx, N. L. Pedersen, C. Pattaro, A. N. Parker, G. Pare, B. A. Oostra, C. J. O'Donnell, M. S. Nieminen, D. A. Nickerson, G. W. Montgomery, T. Meitinger, R. McPherson, M. I. McCarthy, W. McArdle, D. Masson, N. G. Martin, F. Marroni, M. Mangino, P. K. E. Magnusson, G. Lucas, R. Luben, R. J. F. Loos, M.-L. Lokki, G. Lettre, C. Langenberg, L. J. Launer, E. G. Lakatta, R. Laaksonen, K. O. Kyvik, F. Kronenberg, I. R. Konig, K.-T. Khaw, J. Kaprio, L. M. Kaplan, A. Johansson, M.-R. Jarvelin, A. Cecile J. W. Janssens, E. Ingelsson, W. Igl, G. Kees Hovingh, J.-J. Hottenga, A. Hofman, A. A. Hicks, C. Hengstenberg, I. M. Heid, C. Hayward, A. S. Havulinna, N. D. Hastie, T. B. Harris, T. Haritunians, A. S. Hall, U. Gyllensten, C. Guiducci, L. C. Groop, E. Gonzalez, C. Gieger, N. B. Freimer, L. Ferrucci, J. Erdmann, P. Elliott, K. G. Ejebe, A. Doring, A. F. Dominiczak, S. Demissie, P. Deloukas, E. J. C. de Geus, U. de Faire, G. Crawford, F. S. Collins, Y.-d. I. Chen, M. J. Caulfield, H. Campbell, N. P. Burt, L. L. Bonnycastle, D. I. Boomsma, S. M. Boekholdt, R. N. Bergman, I. Barroso, S. Bandinelli, C. M. Ballantyne, T. L. Assimes, T. Quertermous, D. Altshuler, M. Seielstad, T. Y. Wong, E. S. Tai, A. B. Feranil, C. W. Kuzawa, L. S. Adair, H. A. Taylor Jr, I. B. Borecki, S. B. Gabriel, J. G. Wilson, H. Holm, U. Thorsteinsdottir, V. Gudnason, R. M. Krauss, K. L. Mohlke, J. M. Ordovas, P. B. Munroe, J. S. Kooner, A. R. Tall, R. A. Hegele, J. J. P. Kastelein, E. E. Schadt, J. I. Rotter, E. Boerwinkle, D. P. Strachan, V. Mooser, K. Stefansson, M. P. Reilly, N. J. Samani, H. Schunkert, L. A. Cupples, M. S. Sandhu, P. M. Ridker, D. J. Rader, C. M. van Duijn, L. Peltonen, G. R. Abecasis, M. Boehnke, and S. Kathiresan. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707-713.

- Thiele, I., A. Heinken, and R. M. T. Fleming. 2012. A systems biology approach to studying the role of microbes in human health. *Current Opinion in Biotechnology* 24: 4-12.
- Wakil, S. J., and L. A. Abu-Elheiga. 2009. Fatty acid metabolism: Target for metabolic syndrome. *Journal of Lipid Research* 50: S138-S143.
- Wang, F., M. Liu, B. Song, D. Li, H. Pei, Y. Guo, J. Huang, and D. Zhang. 2012a. Prediction and characterization of protein-protein interaction networks in swine. *Proteome Sci* 10: 2.
- Wang, J., J. Jiang, W. Fu, L. Jiang, X. Ding, J.-F. Liu, and Q. Zhang. 2012b. A genome-wide detection of copy number variations using snp genotyping arrays in swine. *BMC Genomics* 13: 273.
- Wang, L., W. Zheng, H. Zhao, and M. Deng. 2013. Statistical analysis reveals co-expression patterns of many pairs of genes in yeast are jointly regulated by interacting loci. *PLoS Genet* 9: e1003414.
- Wang, Z., M. Gerstein, and M. Snyder. 2009. Rna-seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57-63.
- Wood, J., and C. Whittemore. 2007. Pig meat and carcass quality Whittemore's science and practice of pig production. p 4-64. Blackwell Publishing Ltd.
- Wood, J. D., M. Enser, A. V. Fisher, G. R. Nute, P. R. Sheard, R. I. Richardson, S. I. Hughes, and F. M. Whittington. 2008. Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci* 78: 343-358.
- Wright, D., H. Boije, J. R. Meadows, B. Bed'hom, D. Gourichon, A. Vieaud, M. Tixier-Boichard, C. J. Rubin, F. Imsland, F. Hallbook, and L. Andersson. 2009. Copy number variation in intron 1 of *sox5* causes the pea-comb phenotype in chickens. *PLoS Genet* 5: e1000512.

- Xi, R., T.-M. Kim, and P. J. Park. 2010. Detecting structural variations in the human genome using next generation sequencing. *Briefings in Functional Genomics* 9: 405-415.
- Yang, B., A. Bassols, Y. Saco, and M. Perez-Enciso. 2011. Association between plasma metabolites and gene expression profiles in five porcine endocrine tissues. *Genetics Selection Evolution* 43: 28.
- Zhang, W., J. Zhu, E. E. Schadt, and J. S. Liu. 2010. A bayesian partition method for detecting pleiotropic and epistatic eqtl modules. *PLoS Comput Biol* 6: e1000642.
- Zhu, J., P. Y. Lum, J. Lamb, D. GuhaThakurta, S. W. Edwards, R. Thieringer, J. P. Berger, M. S. Wu, J. Thompson, A. B. Sachs, and E. E. Schadt. 2004. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res* 105: 363-374.
- Zhu, J., P. Sova, Q. Xu, K. M. Dombek, E. Y. Xu, H. Vu, Z. Tu, R. B. Brem, R. E. Bumgarner, and E. E. Schadt. 2012. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol* 10: e1001301.
- Zhu, J., M. C. Wiener, C. Zhang, A. Fridman, E. Minch, P. Y. Lum, J. R. Sachs, and E. E. Schadt. 2007. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol* 3: e69.
- Zhu, J., B. Zhang, E. N. Smith, B. Drees, R. B. Brem, L. Kruglyak, R. E. Bumgarner, and E. E. Schadt. 2008. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 40: 854-861.

CHAPTER VII



PAPERS

PAPER I



Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip

Yulixis Ramayo-Caldas, Anna Castelló, Romi N Pena, Estefania Alves, Anna Mercadé,
Carla A Souza, Ana I Fernández, Miguel Perez-Enciso and Josep M Folch

BMC Genomics 2010, **11**:593 doi:10.1186/1471-2164-11-593 published online October 22, 2010

RESEARCH ARTICLE

Open Access

Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip

Yulixaxis Ramayo-Caldas^{1*}, Anna Castelló¹, Romi N Pena^{2,5}, Estefania Alves³, Anna Mercadé¹, Carla A Souza^{1,6}, Ana I Fernández³, Miguel Perez-Enciso^{1,4}, Josep M Folch¹

Abstract

Background: Recent studies in pigs have detected copy number variants (CNVs) using the Comparative Genomic Hybridization technique in arrays designed to cover specific porcine chromosomes. The goal of this study was to identify CNV regions (CNVRs) in swine species based on whole genome SNP genotyping chips.

Results: We used predictions from three different programs (cnvPartition, PennCNV and GADA) to analyze data from the Porcine SNP60 BeadChip. A total of 49 CNVRs were identified in 55 animals from an Iberian x Landrace cross (IBMAP) according to three criteria: detected in at least two animals, contained three or more consecutive SNPs and recalled by at least two programs. Mendelian inheritance of CNVRs was confirmed in animals belonging to several generations of the IBMAP cross. Subsequently, a segregation analysis of these CNVRs was performed in 372 additional animals from the IBMAP cross and its distribution was studied in 133 unrelated pig samples from different geographical origins. Five out of seven analyzed CNVRs were validated by real time quantitative PCR, some of which coincide with well known examples of CNVs conserved across mammalian species.

Conclusions: Our results illustrate the usefulness of Porcine SNP60 BeadChip to detect CNVRs and show that structural variants can not be neglected when studying the genetic variability in this species.

Background

The pig (*Sus scrofa*) is one of the most widespread livestock species and one of the most economically important worldwide. The porcine genome has a total of 18 autosomes plus the X/Y sex chromosome pair; it is similar in size, complexity and chromosomal organization to the human genome. In contrast to SNPs and microsatellites, structural variations have received considerably less attention in pigs. Copy number variants (CNVs) are DNA segments ranging in length from kilobases to several megabases with a variable number of repeats among individuals [1]. Segmental duplications and CNVs have been extensively studied in other organisms [2-7]. Previous studies at genome scale suggest that CNVs comprise 5-12% of the human and ~4% of the dog genome [5,8-10]. CNVs can influence gene expression, affect several metabolic traits and have been

associated with Mendelian and complex genetic disorders [1].

Recent studies in pigs have detected CNVs using the Comparative Genomic Hybridization (CGH) technique in arrays designed to cover specific porcine chromosomes [11,12]. An alternative, cheaper method for CNV detection is based on whole genome SNP genotyping chips [13-15], but it has not been tested yet, to our knowledge, in the swine species. A high-density porcine SNP BeadChip has recently been released by *Illumina*, which contains probes to genotype 62,163 SNPs covering the whole genome. This platform has an average distance between SNPs of 39.61 kb in autosomes and 81.28 kb in chromosome X (based on *Sscrofa9* genome sequence assembly) and is a very valuable resource to study pig genetic variability and the molecular dissection of complex traits of economic importance [16].

The goal of this study was to detect CNV regions (CNVRs) from the Porcine SNP60 BeadChip data on autosomal chromosomes using a pedigree from an Iberian x Landrace (IBMAP) cross and to validate them in a collection of unrelated pigs from different origins.

* Correspondence: yulixaxis.ramayo@uab.es

¹Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
Full list of author information is available at the end of the article

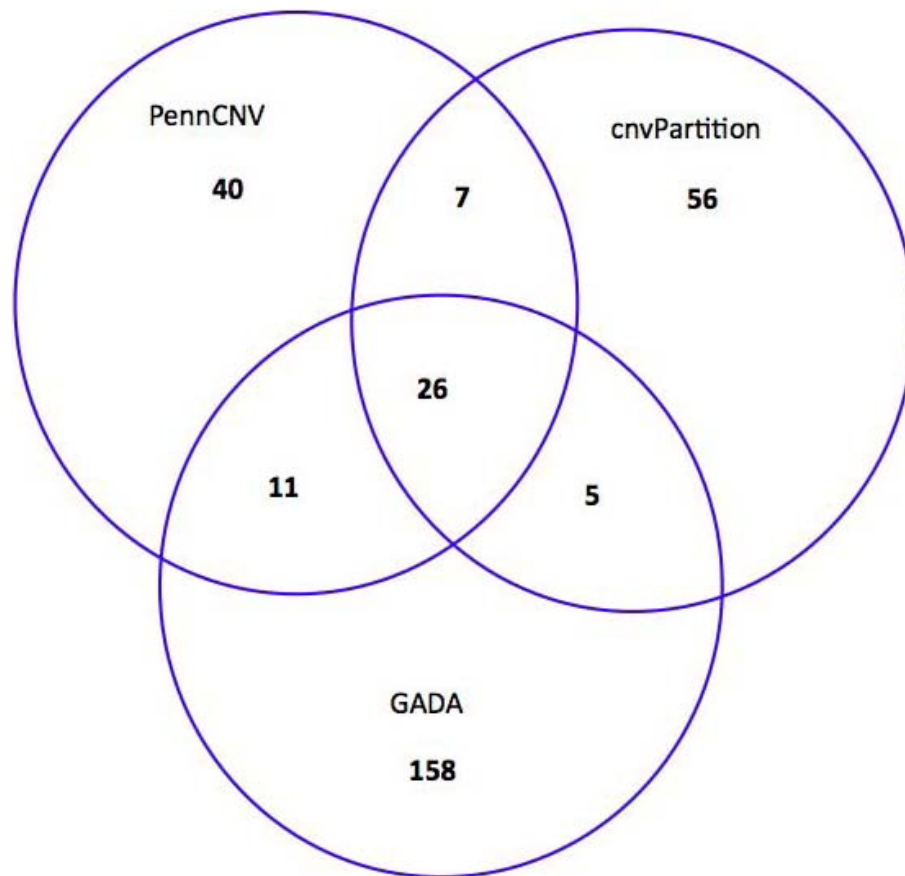


Figure 1 Overlapping CNV events from the three programs used in the analysis.

Results and Discussion

Detection of structural variants

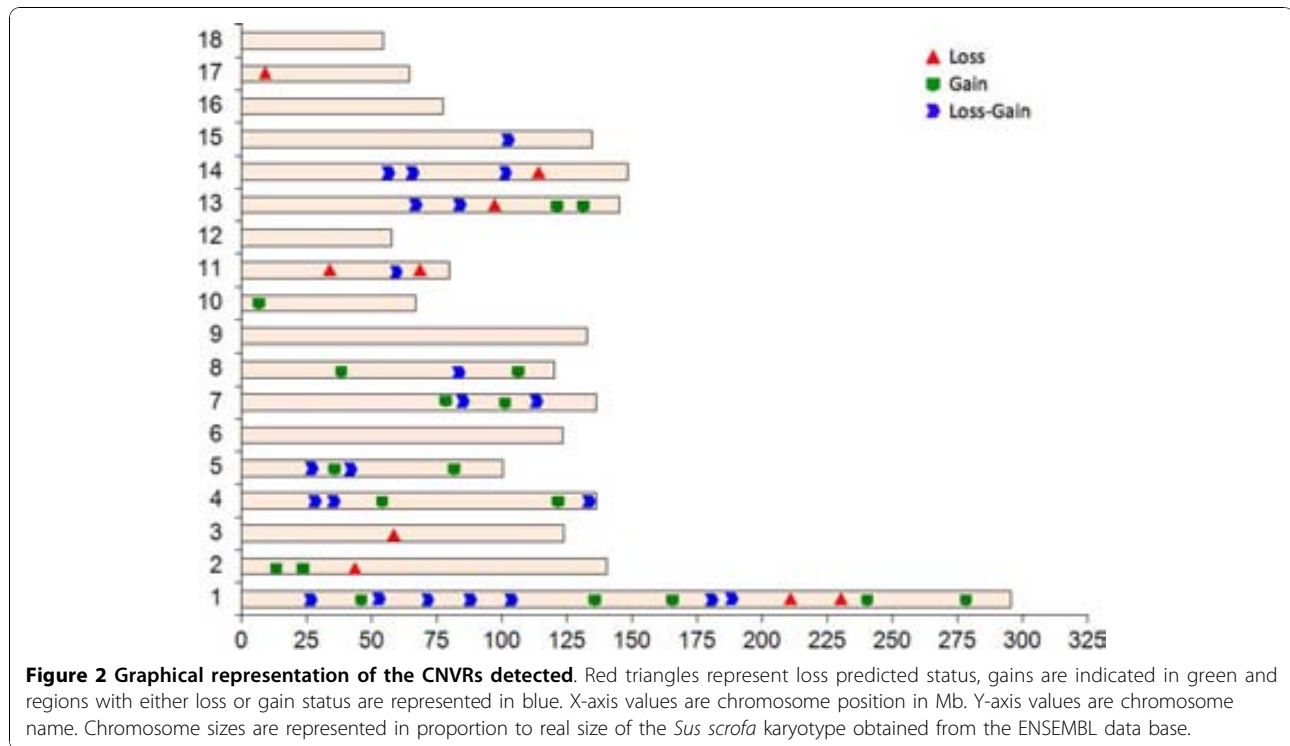
The Porcine SNP60 BeadChip data from 55 IBMAP animals were analyzed by multiple predictions from three different programs: *cnvPartition* (Illumina), *PennCNV* [17] and *GADA* [18]. The initial number of CNVs called by each software was 94, 84, and 200, respectively. Figure 1 summarizes the CNVs identified and compares the results obtained from the three programs.

For further analyses, we retained only CNVs applying a more stringent criterion, namely CNV regions (CNVRs) containing overlapping CNVs recalled by at least two programs, spanning three or more consecutive SNPs and detected in a minimum of two animals. A total of 49 CNVRs located in 13 of the 18 analyzed autosomal chromosomes were identified (Figure 2). All of these CNVRs showed Mendelian inheritance in animals across several generations of the IBMAP cross and therefore are unlikely to be artefacts or false positives, suggesting that our empirical criterion to retain CNVRs is reasonable.

The percentage of CNVRs confirmed by at least two programs was 52.38% for *PennCNV*, 21% for *GADA*

and 40.42% for *cnvPartition*. A total of 26 CNVRs (53.06%) were detected by the three algorithms (Figure 1). Similar results were reported by Winchester *et al.* (2009) comparing different algorithms for CNV detection, suggesting that *PennCNV* is the most accurate program in the prediction of CNVs for the Illumina's platform [19]. In a recent study [20], the relative performance of seven methods for CNV identification was evaluated showing that the *PennCNV* algorithm has a moderate power and the lowest false positive rate. This is likely explained by the unique ability of this algorithm to integrate family relationships and signal intensities from parent-offspring trios data. The low percentage of CNVs called by the *GADA* software might be explained by the relative low coverage of the Porcine SNP60 BeadChip.

The size of the CNVRs detected ranged from 44.7 kb to 10.7 Mb, with a median size of 754.6 kb (Table 1). The Porcine SNP60 BeadChip was originally developed for high-throughput SNP genotyping in association studies. Although CNV detection is feasible with this technology, it is impaired by low marker density, non-uniform distribution of SNPs along pig chromosomes



and lack of non-polymorphic probes specifically designed for CNV identification [16]. Hence, only the largest CNVRs are expected to be assessed with the Porcine SNP60 BeadChip. This explains the difference in minimum CNV length between our study (44.7 kb) and the work of Fadista et al., 2008 (9.3 kb) using the CGH technique.

Among the first 55 animals analyzed, a single CNVR (*CNVR35*) was called in two animals whereas the remaining 48 CNVRs were called in three or more animals. A segregation analysis was performed in 372 additional animals from the IBCMAP cross and the distribution of the CNVRs was additionally studied in 133 unrelated pig samples from different geographical origins (see Methods). All initially detected 49 CNVRs were segregating in the IBCMAP cross and 41 were also detected in American pig populations (Additional file 1, Table S1). The number of animals with alternative alleles for the CNVRs ranged from five (*CNVR13*, *CNVR46*) to 270 (*CNVR15*). The predicted status for the CNVRs was 19 (38.7%) for gain, eight (16.3%) for loss and 22 (45%) for regions with gain or loss status in different animals (Table 1). This proportion may be related to natural selection, as it is assumed that the genome is more tolerant to duplications than to deletions [21-24]. The high percentage of CNVRs with gain or loss status may be the result of including in the analysis pig breeds with different genetic origins and from different countries. However, to establish the real status

of CNVRs, validation by other techniques such as quantitative PCR (qPCR) will be necessary.

Genes located within CNVRs

The Biomart software in the Ensembl Sscrofa9 Database was used to retrieve genes annotated within the genomic regions of CNVRs. A total of 153 protein-coding genes, four miRNA, six miscRNA, three pseudogenes, two rRNA, two snoRNA and nine snRNA were annotated within the 49 CNVRs (Additional file 2, Table S2). Two or more annotated genes were found in 15 CNVRs, whereas one gene only was located in 14 CNVRs. No annotated genes were identified in 20 CNVRs, but this can be due to the incomplete annotation of the Sscrofa9 genome sequence assembly. In contrast to the high number of genes found in this study, it has been suggested that CNVs are located preferably in gene-poor regions [25,26], probably because CNVs present in gene-rich regions may be deleterious and therefore removed by purifying selection [24].

Validation by quantitative PCR

Real time quantitative assays were designed for CNVR validation on seven genomic regions simultaneously detected with the three programs (CNVRs 1, 3, 15, 17, 22, 32, and 36; Table 1). Five of these CNVRs (15, 17, 22, 32, and 36) were confirmed by qPCR, nevertheless fewer animals were validated for CNVRs 15, 17, and 32 (Additional file 3, Fig. S1). Thus, the false discovery rate

Table 1 Description of the 49 CNVRs detected in the pig genome

CNVR ID	Chr.	Start	End	Length (Kb)	Status	Number of Animals
1	1	24217581	24340263	122.68	Loss-Gain	13
2	1	45428275	45549702	121.43	Gain	7
3	1	48467320	48606417	139.10	Loss-Gain	24
4	1	65873816	72839149	6965.33	Loss-Gain	50
5	1	84499444	86580142	2080.70	Loss-Gain	101
6	1	102184641	102246547	61.91	Loss-Gain	73
7	1	160653704	161917767	1264.06	Gain	49
8	1	179815713	179914790	99.08	Loss-Gain	197
9	1	181518754	181741363	222.61	Loss-Gain	86
10	1	206491920	206869145	377.23	Loss	12
11	1	222291368	233007189	10715.82	Loss	11
12	1	237242221	237929823	687.60	Gain	8
13	2	11601476	11714214	112.74	Gain	5
14	2	21970642	22174716	204.07	Gain	7
15	2	40533655	41466383	932.73	Loss	270
16	3	56202930	56347970	145.04	Gain	9
17	4	24971805	25077329	105.52	Loss-Gain	39
18	4	33537533	33962195	424.66	Loss-Gain	8
19	4	50569393	51322987	753.59	Gain	14
20	4	119210502	119256548	46.05	Gain	7
21	4	134367793	134519459	151.67	Loss-Gain	21
22	5	23899971	24070933	170.96	Loss-Gain	51
23	5	35733150	37322403	1589.25	Loss-Gain	51
24	5	35933178	36136940	203.76	Gain	12
25	7	77371678	77536074	164.40	Gain	18
26	7	82665585	83053927	388.34	Loss-Gain	18
27	7	97896821	98115996	219.18	Gain	21
28	7	110105319	110156658	51.34	Loss-Gain	11
29	10	4479233	4701713	222.48	Gain	42
30	11	65309203	65381195	71.99	Loss	20
31	11	56381032	57812846	1431.81	Loss-Gain	45
32	13	64352825	64798051	445.23	Loss-Gain	105
33	13	118821556	118923746	102.19	Gain	26
34	14	53301700	55310453	2008.75	Loss-Gain	10
35	14	63834660	63882223	47.56	Loss-Gain	176
36	14	111363926	111540634	176.71	Loss	150
37	14	97526178	99696847	2170.67	Loss-Gain	49
38	15	99843606	100116097	272.49	Loss-Gain	75
39	17	6525429	6635237	109.81	Loss	84
40	8	80936785	81030481	93.70	Loss-Gain	93
41	13	81503361	81601546	98.19	Loss-Gain	118
42	1	130025060	130085466	60.41	Gain	26
43	1	276666775	276786004	119.23	Gain	13
44	5	79779580	79931050	151.47	Gain	34
45	8	35691571	36104826	413.26	Gain	53
46	8	104917016	104968103	51.09	Gain	5
47	11	30072771	30117419	44.65	Loss	38
48	13	94769297	94922644	153.35	Loss	66
49	13	128150975	128360061	209.09	Gain	17

The genomic coordinates are expressed in bp and are relative to the *Sus scrofa* April 2009 genome sequence assembly (Sscrofa9)

(FDR) for the seven analyzed CNVRs was 29%; it should be noted that the percentage of CNVRs validated in this study (71%) is higher than previously reported in pigs (50%) [11]. This result might be explained by the stringent criteria used in our analysis, which was proposed in order to increase confidence and minimize the false positives. Nevertheless, we were not able to confirm two of the CNVRs. Several factors may account for the discrepancy in CNVR prediction between the *in silico* analysis of Porcine SNP60 BeadChip data and the qPCR method. First, the incomplete status of the 4× sequence depth Sscrofa9 assembly and the low probe density of the Porcine SNP60 BeadChip makes it difficult to establish the true boundaries of CNVRs and may result in an over estimation of their real size. Therefore, it cannot be ruled out that the primers used to validate the CNVRs by qPCR may have been designed outside the structural polymorphic region. Second, polymorphisms such as SNPs and indels may influence the hybridization of the qPCR primers, changing the relative quantification (RQ) values for some animals. Finally, the true CNVR boundaries may be also polymorphic between the analyzed animals.

For the qPCR validation of *CNVR36*, a PCR protocol for the *Cytochrome P450 C32 Fragment gene* [EMBL: ENSSSCG00000010487] was designed. A total of 37 animals were analyzed: 21 with statistical evidence for CNVR and 16 without the CNVR (control group). One of the animals from the control group was selected as reference. Six false positive animals were observed, indicating a FDR of 29% for *CNVR36* (Figure 3).

A qPCR assay with primers located in the *SLC16A7* gene [EMBL: ENSSSCG00000000456] was used for *CNVR22* validation. A total of 50 animals were analyzed: 21 with statistical evidence for CNVR (12 from the IBSMAP cross and nine unrelated individuals belonging to six different breeds of American populations) and 29 without the CNVR (control group). One of the animals from the control group was selected as reference. Nine of the IBSMAP cross animals were validated by qPCR (FDR = 25%). Conversely, only three animals from the American populations were validated by qPCR, suggesting a higher FDR (67%) (Figure 4). These differences in FDR may be explained by the higher accuracy of the PennCNV algorithm when family information is available and stress the usefulness of including family information in CNV detection. However, this conclusion should be taken with caution due to the limited number of animals analyzed.

For CNVRs 22 and 36, copy number changes were also identified by qPCR in animals where CNVs were not detected initially in the statistical analysis (three and eight animals, respectively). This represents a false negative rate of 10% (3/29) for *CNVR22* and 50% (8/16) for

CNVR36. The three false negative animals for *CNVR22* were classified as deletions by qPCR protocol. A similar situation, but with a different copy number status, was observed for *CNVR36*, where the eight false negative animals showed a duplication pattern by qPCR. False negative identification is common in CNV detection, and has been reported previously using the CGH technique in pigs and other mammalian species [5,11].

Three of the validated CNVRs (17, 22, and 36) showed differential patterns of copy number variants between breeds. For instance, *CNVR22* showed a loss (deletion) in Landrace and in animals from other breeds (Figure 4). Assuming that Iberian pigs have two copies of *CNVR22* (qPCR RQ = 1), five animals showing an RQ = 0 by qPCR are predicted to be homozygous for a deletion on this genomic region. In *CNVR36*, a loss was found in Iberian pigs relative to Landrace animals (Figure 3). In agreement with the Mendelian segregation of this CNVR, hybrid animals show intermediate RQ values. The RQ mean values were 0.49 for Iberian, 2.51 for Landrace and 1.2 for hybrid animals.

CNVR36 contains a miRNA gene [EMBL: ENSSSCG00000019484] and the *Cytochrome P450 C32 Fragment gene* (Additional file 2, Table S2), which is a member of the *Cytochrome P450* gene family (*CYTP450*). Proteins coded by this gene family constitute the major catalytic component of the liver mixed-function oxidase system and play a pivotal role in the metabolism of many endogenous and exogenous compounds [27]. Interestingly, CNVs comprising genes of the *CYTP450* family have been described in humans and dogs [5,10,28], but had not been previously reported in pigs. In humans, copy number variations of *CYTP450* genes have been associated with variation in drug metabolism phenotypes [29-31]. Differential expression of genes of the *CYTP450* family has been correlated with androsterone levels in pigs from Duroc and Landrace breeds [32]. It has also been demonstrated that the total *CYTP450* activity was slightly higher in minipigs compared to conventional pigs [33]. *CNVR36* lays close to the peak position of a QTL for androsterone levels described in a cross between Large White and Chinese Meishan [34]. This suggests a possible role of this structural variation in determining androsterone levels; however, more studies will be necessary to validate this hypothesis.

CNVR22, also validated by qPCR, comprises the *SLC16A7* gene. This gene belongs to the *solute carrier family 16* gene family, which encodes 14 proteins that are largely known as monocarboxylate transporters (MCTs). The human *SLC16A7* gene encodes the MCT2 protein [35] and it is expressed in several normal human tissues. In pigs, MCT2 may function as a house-keeping lactate transporter, regulating the acidification

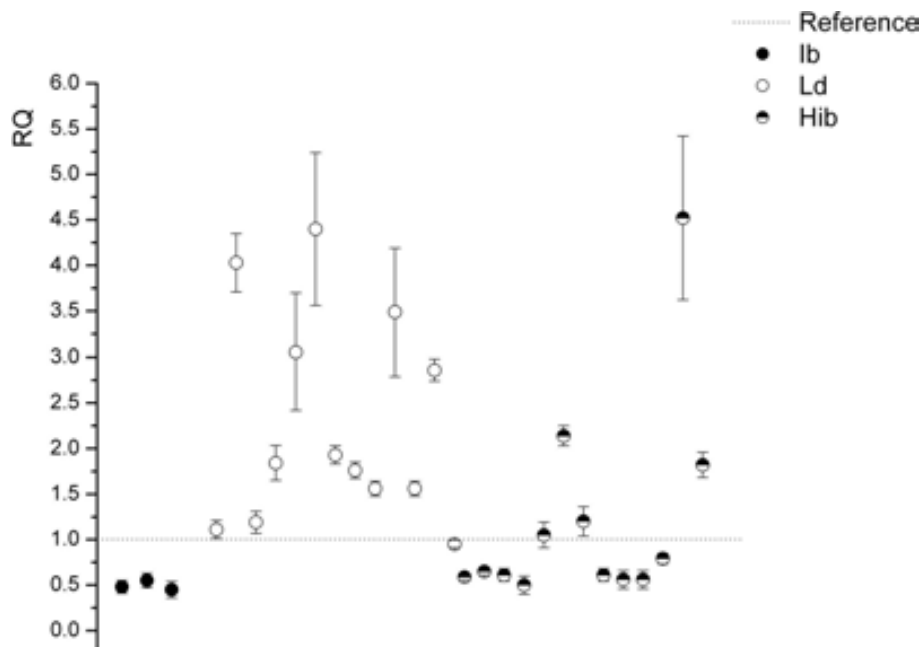


Figure 3 Analysis by quantitative PCR (qPCR) of *CNVR36* (*CYP4502 C32 Fragment gene*). Twenty-one animals with statistical evidence for CNVR and eight false negative animals from the control group are shown. The horizontal dashed line represents the relative quantification (RQ) value of the reference animal. Each dot represents the relative copy number in comparison to the reference individual. Y-axis shows the RQ value obtained by qPCR. Vertical bars represent the standard error. Breed abbreviations are: **Ib**: Iberian; **Ld**: Landrace; **Hib**: animals belonging to several generations of the IBCMAP cross (F1, F2, and BC); **Mx**: Mexican hairless; **Brz**: Brazilian local breed; **Gu**: Guatemala local breed; **Yu**: Yucatan miniature pig, **CC**: Cuban creole pig.

of glycolytic muscles [36]. Remarkably, *CNVR22* is located in the middle of the confidence interval of a QTL for meat pH described in four pig populations [37].

Duplication events have also been validated by qPCR for *SOX14* [EMBL: ENSSSCG00000011656] (*CNVR32*) and *INSC* [EMBL: ENSSSCG00000013385] (*CNVR15*). Copy number changes have not been previously reported in either of them in pigs. *SOX14* is a member of the *SOX* gene family [38] of transcription factors

involved in the regulation of embryo development and cell fate determination. *SOX14* may have a major role in the regulation of nervous system development and it is a mediator of the neuronal death process [39]. *SOX14* is an intronless gene that may have arisen by duplication from an ancestral *SOX B* gene, which likely was the product of a retrotransposition event [40]. *Inscuteable* (*INSC*) was first described in *Drosophila* and it plays a central role in the molecular machinery for mitotic spindle orientation and regulates cell polarity for asymmetric

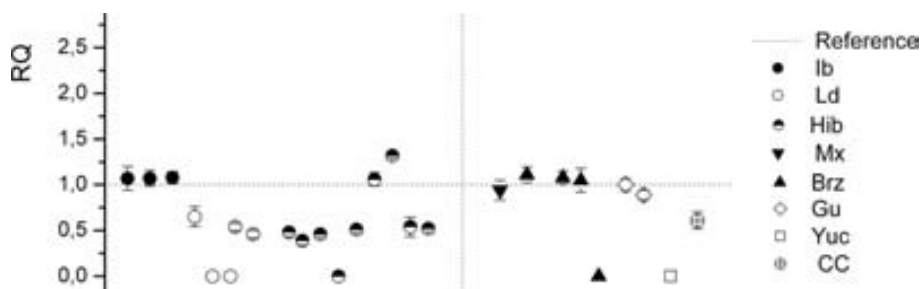


Figure 4 Analysis by quantitative PCR (qPCR) of *CNVR22* (*SLC16A7 gene*). Twenty-one animals with statistical evidence for CNVR, three false negative and two Iberian from the control group are plotted. The horizontal dashed line represents the relative quantification (RQ) value of the reference animal. Each dot represents the relative copy number of the animal in comparison to the reference. Y-axis shows the RQ value obtained by qPCR. Vertical bars represent the standard error. The vertical dashed line separates the 17 related IBCMAP individuals from the nine unrelated American local breeds. Breed abbreviations as described in Figure 3.

division [41,42]. *Inscuteable homologs* have been found in several species, including vertebrates and insects [43]. In mammals, INSC is functionally conserved and it is required for correct orientation of the mitotic spindle in retina [43] and skin [44] precursor cells.

The qPCR assay for *CNVR17* validation was designed over the sequence of one expressed sequence tag [EMBL: EW037329]. From four Cuban creole pigs tested, three animals showed a deletion and one animal a duplication event (Additional file 3, Fig. S1).

Other relevant CNVRs

Although other CNVRs have not been analyzed by qPCR, there is evidence of structural polymorphism in the literature. For instance, *CNVR45* contains the *KIT* gene, a well-characterized and functionally important CNV in pigs. The dominant white coat phenotype in pigs is caused by *KIT* gene duplication or triplication and a splice mutation in one of the *KIT* gene copies [45-49]. In addition, studies in other mammals [5,6,50-55] have described CNVRs overlapping other gene families including: *Olfactory receptor family*, *Glutamate receptor family*, *Solute carrier family*, *Cytochrome P450 family*, *Cyclic nucleotide phosphodiesterases family* and *Fucosyltransferase family*. Twelve of the CNVRs detected in our study include or overlap porcine orthologues of these genes. Furthermore, 13 of the detected CNVRs include 47 genes previously reported in the Human Database of Genomic Variants <http://projects.tcag.ca/variation/?source=hg19> [56] (Additional file 4, Table S3).

Conclusions

We have described the first CNVRs in swine based on whole genome SNP genotyping chips. A total of 49 CNVRs were identified in 13 autosomal chromosomes. These CNVRs showed Mendelian inheritance across 427 individuals belonging to several generations of an Iberian x Landrace cross, and were also confirmed in different pig breeds. Five out of seven selected CNVRs were validated by qPCR; among the remaining CNVRs we found well known examples of CNVs conserved across mammalian species. Although these results illustrate the usefulness of Porcine SNP60 BeadChip to detect CNVRs, the number detected here is probably a gross underestimate given the wide interval between SNPs in the Porcine 60 k BeadChip.

Methods

Animal Material

We analyzed a total of 560 animals, including 427 individuals (150 males and 277 females) belonging to several generations of the IBSMAP cross. This population was originated by crossing three Iberian (Guadyrvas line)

boars with 31 Landrace sows [57,58] (Additional file 5, Fig. S2). The remaining 133 pig samples were obtained from different geographical origins: 127 from American local breeds and village pigs [59], four black Sicilian pigs, one Hungarian Mangalitzka and one Chinese Wild boar (Additional file 6, Table S4). We adhered to our national and institutional guidelines for the ethical use and treatment of animals in experiments.

Genotyping

All 560 animals were genotyped with the Porcine SNP60 BeadChip (Illumina Inc., USA) using the *Infinium HD Assay Ultra* protocol (Illumina). Raw data had high-genotyping quality (call rate >0.99) and were visualized and analyzed with the GenomeStudio software (Illumina). For subsequent data analysis, a subset of 50,572 SNPs was selected by removing the SNPs located in sex chromosomes and those not mapped in the Sscrofa9 assembly.

Statistical analysis

Following the recommendations of Winchester et al. (2009) to increase the confidence in CNV detection and limit the number of false positives, we used predictions from multiple programs. First, we used the Illumina's proprietary software GenomeStudio to check data quality and the *cnvPartition v2.4.4 Analysis Plug-in* for CNV detection. The minimum probe count employed was three and the remaining parameters were used according to the default criteria provided (Illumina). Then, we exported the signal intensity data of $\log R_{ratio}$ and B allele frequency to employ the R package for Genome Alteration Detection Algorithm (GADA) [18], which includes one algorithm based in sparse Bayesian learning to predict CNV changes. The multiple array analysis option was employed and the parameters defined for the Bayesian learning model and the backward elimination (BE) were: 0.8 for sparseness hyperparameter (α), 8 for critical value of the BE and 3 as the minimum number of SNPs at each segment.

Next, we used the command line version of PennCNV software that integrates, in a joint-calling algorithm, a Hidden Markov Model (HMM) with family relationships, signal intensities for parent-offspring trios, marker distance and population frequency of allele B [17]. The CNV calling was performed using the default parameters of the HMM model with 0.01 of UF factor. The "-trio" and "-quartet" arguments were employed to make use of our family information.

It is unclear with this kind of data, where the statistical properties of the methods are unknown, which is the optimum strategy to balance false positives and power. Here we chose to follow a pragmatic approach, requiring that the CNV was called by at least two algorithms,

detected in at least two animals and contained three or more consecutive SNPs. Hence, these genomic regions should be referred as copy number variable regions (CNVRs). To define the size of each CNVR in the genome, we used the overlapping region between CNV predictions from different programs.

Pipeline analysis for CNVR detection was initially performed in 55 individuals of the IBSMAP cross (13 males and 42 females), including all founder Iberian boars (three males), 24 founder Landrace sows, 17 F1, three F2, and eight backcross animals. Subsequently, we tested the segregation of these initially detected CNVRs in the rest of the IBSMAP cross animals (372), and described their distribution in 127 unrelated pig samples from American local pigs, four black Sicilian pigs, one Hungarian Mangalitza and one Chinese Wild boar.

Gene annotation within the CNVRs was retrieved from the Ensembl Genes 57 Database using the Biomart [http://www.biomart.org] software.

Quantitative real time PCR

Quantitative real time PCR (qPCR) was used to validate seven genomic regions detected by the three methods and representing different predicted status of copy numbers. We used the $2^{-\Delta\Delta C_t}$ method for relative quantification (RQ) of CNVs [5,60,61]. This comparative method uses a target assay for the DNA segment being interrogated for copy number variation and a reference assay for an internal control segment, which is normally a known single copy gene; moreover a reference sample is included. The method requires the target and reference PCR efficiencies to be nearly to equal. Experiments were performed on the test and control primers to verify comparable efficiency in amplification prior to analysis of copy number.

CNVRs were quantified using the Taqman chemistry in an ABI PRISM[®] 7900HT instrument (Applied Biosystems, Inc., Foster City, CA); results were analyzed with the SDS software (Applied Biosystems). Primers and hydrolysis probes (Taqman-MGB labeled with FAM) were designed for the seven CNVR regions with the Primer Express software (Applied Biosystems). A previously described [62] design on the *glucagon* gene [EMBL:GCG] was used as single copy control region, but a single nucleotide substitution on primer forward was introduced to adapt the primer to the porcine species. Primers and probes are shown in Additional file 7, Table S5.

PCR amplifications were performed in a total volume of 20 μ l containing 10 ng of genomic DNA. Taqman PCR Universal Master Mix (Applied Biosystems) was used in all reactions except in GCG amplifications, where TaqMan[®] PCR Core Reagents (Applied Biosystems) with 2.5 mM MgCl₂ were utilized. All primers

and probes were used at 900 nM and 250 nM respectively, except *CytochromeP450 2C32 Fragment* forward primer, which was used at 300 nM. Each sample was analyzed in triplicate. The thermal cycle was: 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 sec at 95°C and 1 min at 60°C. One sample without copy number variation for each of the genomic regions analyzed was used as reference.

Data availability

The full data set have been submitted to dbVAR [63] under the accession number nstd44.

Additional material

Additional file 1: Table S1. Results of the distribution analysis in American pig populations.

Additional file 2: Table S2. Gene annotation within the CNVRs retrieved from the Ensembl Genes 57 Database using the Biomart software.

Additional file 3: Fig. S1. Results of quantitative PCR (qPCR) for CNVRs 15 (top), 17 (middle), and 32 (bottom). A total of 17 animals are showed in each plot. Breed abbreviations are: **Ib**: Iberian; **Ld**: Landrace; **Hib**: animals belonging to several generations of the IBSMAP cross (F1, F2, and BC); **CC**: Cuban creole pig; **Gu**: Guatemala local breed; **Yu**: Yucatan miniature pig; **Pe**: Peruvian creole pig.

Additional file 4: Table S3. List of pig genes previously reported in the Human Database of Genomic Variants.

Additional file 5: Fig. S2. Structure of the IBSMAP cross. Abbreviations are: **Ib**: Iberian; **Ld**: Landrace; **F1**: first generation; **F2**: second generation; **F3**: third generation; **BC**: first backcross; **BC1_LD**: second backcross.

Additional file 6: Table S4. Description of samples from American local breeds.

Additional file 7: Table S5. Primers and probes used in quantitative PCR validation

Abbreviations

CNV: copy number variation; **CNVR**: CNV region; **PCR**: polymerase chain reaction; **IBSMAP**: Iberian x Landrace intercross; **qPCR**: quantitative real time PCR; **RQ**: relative quantification value; **CYP450**: *Cytochrome P450* gene family; **SLC16A7**: solute carrier family 16 member 7; **MCT2**: monocarboxylic acid transporter 2; **SOX14**: SRY (sex determining region Y)-box 14; **INSC**: inscuteable homolog (*Drosophila*).

Acknowledgements

This work was funded by MICINN project AGL2008-04818-C03/GAN (Spanish Ministry of Science) and by Innovation Consolider-Ingenio 2010 Programme (CSD2007-00036 "Centre for Research in Agrigenomics"). Y. Ramayo-Caldas was funded by a FPU PhD grant from Spanish Ministerio de Educación (AP2008-01450). C.A. Souza was funded by a PhD grant from CAPES, Brazil. We thank S. Scherrer, R. Pique-Regi, Yang Bin and A. Esteve for the help with data analysis. We also thank Martien Groenen (Wageningen, NL) for providing information about SNP positioning in Assembly 9.

Author details

¹Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. ²Genètica i Millora Animal, IRTA Lleida, 25198 Lleida, Spain. ³Departamento de Mejora Animal SGIT-INIA, 28040 Madrid, Spain. ⁴Institut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. ⁵Department of Animal Production, University of Lleida, 25198 Lleida, Spain. ⁶Universidade Católica de Brasília, Brazil.

Authors' contributions

YRC, AIF, MPE and JMF conceived and designed the experiment. JMF was the principal investigator of the project. YRC performed the data analysis and drafted the manuscript. EA, RNP, AIF, CAS, YRC, MPE and JMF collected samples. RNP and CAS performed DNA isolation. AM and AC performed the SNP genotyping and AC did the qPCR and RT-PCR assays. All authors read and approved the final manuscript.

Received: 15 June 2010 Accepted: 22 October 2010

Published: 22 October 2010

References

- Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ: **Copy number variation influences gene expression and metabolic traits in mice.** *Hum Mol Genet* 2009, **18**:4118-4129.
- Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim JI, Seo JS, Yakhini Z, Laderman S, Bruhn L, Lee C: **The fine-scale and complex architecture of human copy-number variation.** *Am J Hum Genet* 2008, **82**:685-695.
- Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD, Hubner N, Cuppen E: **Distribution and functional impact of DNA copy number variation in the rat.** *Nat Genet* 2008, **40**:538-545.
- She X, Cheng Z, Zollner S, Church DM, Eichler EE: **Mouse segmental duplication and copy number variation.** *Nat Genet* 2008, **40**:909-914.
- Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM: **The genomic architecture of segmental duplications and associated copy number variants in dogs.** *Genome Res* 2009, **19**:491-499.
- Bae JS, Cheong HS, Kim LH, Namgung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD: **Identification of copy number variations and common deletion polymorphisms in cattle.** *BMC Genomics* 2010, **11**:232.
- Wang X, Nahashon S, Feaster T, Bohannon-Stewart A, Adefope N: **An initial map of chromosomal segmental copy number variations in the chicken.** *BMC Genomics* 2010, **11**:351.
- Li J, Yang T, Wang L, Yan H, Zhang Y, Guo Y, Pan F, Zhang Z, Peng Y, Zhou Q, He L, Zhu X, Deng H, Levy S, Papiasian CJ, Drees BM, Hamilton JJ, Recker RR, Cheng J, Deng HW: **Whole Genome Distribution and Ethnic Differentiation of Copy Number Variation in Caucasian and Asian Populations.** *PLoS One* 2009, **4**:e7958.
- Choy KW, Setlur SR, Lee C, Lau TK: **The impact of human copy number variation on a new era of genetic testing.** *BJOG* .
- Yim SH, Kim TM, Hu HJ, Kim JH, Kim BJ, Lee JY, Han BG, Shin SH, Jung SH, Chung YJ: **Copy number variations in East-Asian population and their evolutionary and functional implications.** *Hum Mol Genet* 19:1001-1008.
- Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C: **A snapshot of CNVs in the pig genome.** *PLoS One* 2008, **3**:e3916.
- Tang H, Li F, Finlayson HA, Smith S, Lu Z, Langford C, Archibald AL: **Structural And Copy Number Variation In The Pig Genome.** In *Book Structural And Copy Number Variation In The Pig Genome*. Edited by: City: Plant & Animal Genomes XVIII Conference. Town 2010; January 9-13, 2010.
- Tuefferd M, Bondt AD, Wyngaert IVD, Talloen W, Verbeke T, Carvalho B, Clevert DA, Alifano M, Raghavan N, Amaratunga D, Göhlmann H, Broët P, Camilleri-Broët S: **Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays.** *Genes, Chromosomes and Cancer* 2008, **47**:957-964.
- Komura D, Shen F, Ishikawa S, Fitch KR, Chen W, Zhang J, Liu G, Ihara S, Nakamura H, Hurler ME, Lee C, Scherer SW, Jones KW, Shapero MH, Huang J, Aburatani H: **Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays.** *Genome Res* 2006, **16**:1575-1584.
- Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, Cheung SW, Shen RM, Barker DL, Gunderson KL: **High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping.** *Genome Research* 2006, **16**:1136-1148.
- Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MAM: **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.
- Wang K, Li M, Hadley D, Liu R, Glennier J, Grant SF, Hakonarson H, Bucan M: **PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.** *Genome Res* 2007, **17**:1665-1674.
- Pique-Regi R, Monso-Varona J, Ortega A, Seeger RC, Triche TJ, Asgharzadeh S: **Sparse representation and Bayesian detection of genome copy number alterations from microarray data.** *Bioinformatics* 2008, **24**:309-318.
- Winchester L, Yau C, Ragoussis J: **Comparing CNV detection methods for SNP arrays.** *Briefings in Functional Genomics and Proteomics* 2009, **8**:353-366.
- Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ: **Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays.** *Nucl Acids Res* 2010, gkq040.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, et al: **Global variation in copy number in the human genome.** *Nature* 2006, **444**:444-454.
- Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE: **Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome.** *Am J Hum Genet* 2006, **79**:275-290.
- Brewer C, Holloway S, Zawalynski P, Schinzel A, FitzPatrick D: **A chromosomal duplication map of malformations: regions of suspected haplo- and triplolethality--and tolerance of segmental aneuploidy--in humans.** *Am J Hum Genet* 1999, **64**:1702-1708.
- Conrad DF, Hurler ME: **The population genetics of structural variation.** *Nat Genet* 2007.
- Conrad DF, Andrews TD, Carter NP, Hurler ME, Pritchard JK: **A high-resolution survey of deletion polymorphism in the human genome.** *Nat Genet* 2006, **38**:75-81.
- Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurler ME, Carter NP, Scherer SW, Lee C: **Copy number variation: New insights in genome diversity.** *Genome Research* 2006, **16**:949-961.
- Anzenbacher P, Anzenbacherova E: **Cytochromes P450 and metabolism of xenobiotics.** *Cell Mol Life Sci* 2001, **58**:737-747.
- Stankiewicz P, Lupski JR: **Genome architecture, rearrangements and genomic disorders.** *Trends in Genetics* 2002, **18**:74-82.
- Daly AK: **Pharmacogenetics of the cytochromes P450.** *Curr Top Med Chem* 2004, **4**:1733-1744.
- Ledesma MC, Agundez JA: **Identification of subtypes of CYP2D gene rearrangements among carriers of CYP2D6 gene deletion and duplication.** *Clin Chem* 2005, **51**:939-943.
- Ouahchi K, Lindeman N, Lee C: **Copy number variants and pharmacogenomics.** *Pharmacogenomics* 2006, **7**:25-29.
- Grindflek E, Berget I, Moe M, Oeth P, Lien S: **Transcript profiling of candidate genes in testis of pigs exhibiting large differences in androstenone levels.** *BMC Genet* 11:4.
- Skaaniid MT, Friis C: **Cytochrome P450 sex differences in minipigs and conventional pigs.** *Pharmacol Toxicol* 1999, **85**:174-180.
- Lee GJ, Archibald AL, Law AS, Lloyd S, Wood J, Haley CS: **Detection of quantitative trait loci for androstenone, skatole and boar taint in a cross between Large White and Meishan pigs.** *Anim Genet* 2005, **36**:14-22.
- Lin RY, Vera JC, Chaganti RS, Golde DW: **Human monocarboxylate transporter 2 (MCT2) is a high affinity pyruvate transporter.** *J Biol Chem* 1998, **273**:28959-28965.
- Sepponen K, Koho N, Puolanne E, Ruusunen M, Poso AR: **Distribution of monocarboxylate transporter isoforms MCT1, MCT2 and MCT4 in porcine muscles.** *Acta Physiol Scand* 2003, **177**:79-86.
- Srikanthai T, Murani E, Wimmers K, Ponsuksili S: **Four loci differentially expressed in muscle tissue depending on water-holding capacity are associated with meat quality in commercial pig herds.** *Mol Biol Rep* 2009, **37**:595-601.

38. Arsic N, Rajic T, Stanojic S, Goodfellow PN, Stevanovic M: **Characterisation and mapping of the human SOX14 gene.** *Cytogenet Cell Genet* 1998, **83**:139-146.
39. Osterloh JM, Freeman MR: **Neuronal death or dismemberment mediated by Sox14.** *Nat Neurosci* 2009, **12**:1479-1480.
40. Kirby PJ, Waters PD, Delbridge M, Svartman M, Stewart AN, Nagai K, Graves JA: **Cloning and mapping of platypus SOX2 and SOX14: insights into SOX group B evolution.** *Cytogenet Genome Res* 2002, **98**:96-100.
41. Kraut R, Chia W, Jan LY, Jan YN, Knoblich JA: **Role of inscuteable in orienting asymmetric cell divisions in Drosophila.** *Nature* 1996, **383**:50-55.
42. Chia W, Yang X: **Asymmetric division of Drosophila neural progenitors.** *Curr Opin Genet Dev* 2002, **12**:459-464.
43. Zigman M, Cayouette M, Charalambous C, Schleiffer A, Hoeller O, Dunican D, McCudden CR, Firmberg N, Barres BA, Siderovski DP, Knoblich JA: **Mammalian inscuteable regulates spindle orientation and cell fate in the developing retina.** *Neuron* 2005, **48**:539-545.
44. Lechler T, Fuchs E: **Asymmetric cell divisions promote stratification and differentiation of mammalian skin.** *Nature* 2005, **437**:275-280.
45. Moller M, Chaudhary R, Hellmén E, Høyheim B, Chowdhary B, Andersson L: **Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor.** *Mammalian Genome* 1996, **7**:822-830.
46. Marklund S, Kijas J, Rodriguez-Martinez H, Rännstrand L, Funa K, Moller M, Lange D, Edfors-Lilja I, Andersson L: **Molecular Basis for the Dominant White Phenotype in the Domestic Pig.** *Genome Research* 1998, **8**:826-833.
47. Johansson Moller M, Chaudhary R, Hellmen E, Hoyheim B, Chowdhary B, Andersson L: **Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor.** *Mamm Genome* 1996, **7**:822-830.
48. Seo BY, Park EW, Ahn SJ, Lee SH, Kim JH, Im HT, Lee JH, Cho IC, Kong IK, Jeon JT: **An accurate method for quantifying and analyzing copy number variation in porcine KIT by an oligonucleotide ligation assay.** *BMC Genetics* 2007, **8**:81.
49. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A: **Worldwide Phylogeography of Wild Boar Reveals Multiple Centers of Pig Domestication.** *Science* 2005, **307**:1618-1621.
50. Niimura Y, Nei M: **Extensive gains and losses of olfactory receptor genes in mammalian evolution.** *PLoS One* 2007, **2**:e708.
51. Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ: **Different evolutionary processes shaped the mouse and human olfactory receptor gene families.** *Hum Mol Genet* 2002, **11**:535-546.
52. Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, Morin E, Retout E, Valin AS, Lindblad-Toh K, Nicolas J, Galibert F: **The dog and rat olfactory receptor repertoires.** *Genome Biol* 2005, **6**:R83.
53. Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ: **Extensive copy-number variation of the human olfactory receptor gene family.** *Am J Hum Genet* 2008, **83**:228-242.
54. Poot M, Eleveld MJ, Van 't Slot R, Ploos van Amstel HK, Hochstenbach R: **Recurrent copy number changes in mentally retarded children harbour genes involved in cellular localization and the glutamate receptor complex.** *Eur J Hum Genet* 2009, **18**:39-46.
55. Nozawa M, Kawahara Y, Nei M: **Genomic drift and copy number variation of sensory receptor genes in humans.** *Proc Natl Acad Sci USA* 2007, **104**:20421-20426.
56. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**:949-951.
57. Clop A, Ovilo C, Perez-Enciso M, Cercos A, Tomas A, Fernandez A, Coll A, Folch JM, Barragan C, Diaz I, Oliver MA, Varona L, Sillio L, Sanchez A, Noguera JL: **Detection of QTL affecting fatty acid composition in the pig.** *Mamm Genome* 2003, **14**:650-656.
58. Perez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, Folch JM, Babot D, Estany J, Oliver MA, Diaz I, Sanchez A: **A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross.** *J Anim Sci* 2000, **78**:2525-2531.
59. Souza CA, Ramayo Y, Megens HJ, Rodríguez MC, Loarca A, Caal E, Soto H, Melo M, Revidatti MA, de la Rosa SA, Shemereteva IN, Okumura N, Cho IC, Delgado JV, Paiva SR, Crooijmans RPMA, Schook LB, Groenen MAM, Ramos-Onsins SE, Pérez-Enciso M: **Porcine Colonization Of The Americas: A 60 k SNP Story.** *World Congress on Genetics Applied to Livestock Production.* Leipzig, Germany. August 1-6; 2010.
60. 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**:402-408.
61. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ: **A High-Resolution Map of Segmental DNA Copy Number Variation in the Mouse Genome.** *PLoS Genet* 2007, **3**:e3.
62. Ballester M, Castello A, Ibanez E, Sanchez A, Folch JM: **Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals.** *Biotechniques* 2004, **37**:610-613.
63. NCBI: **Database of Genomic Structural Variation.** *Book Database of Genomic Structural Variation* City:[http://www.ncbi.nlm.nih.gov/dbvar].

doi:10.1186/1471-2164-11-593

Cite this article as: Ramayo-Caldas et al.: Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics* 2010 **11**:593.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



PAPER II

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Genome-wide association study for intramuscular fatty acid composition in an Iberian x Landrace cross

Y. Ramayo-Caldas, A. Mercadé, A. Castelló, B. Yang, C. Rodríguez, E. Alves, I. Díaz,
N. Ibáñez-Escriche, J.L. Noguera, M. Pérez-Enciso, A.I. Fernández and J.M. Folch

J ANIM SCI 2012, 90:2883-2893. doi: 10.2527/jas.2011-4900 published online July 10, 2012

**Genome-wide association study for intramuscular fatty acid composition in an Iberian
× Landrace cross**

Y. Ramayo-Caldas, A. Mercadé, A. Castelló, B. Yang, C. Rodríguez, E. Alves, I. Díaz, N. Ibáñez-Escriche, J.L. Noguera, M. Pérez-Enciso, A.I. Fernández and J.M. Folch

J ANIM SCI 2012, 90:2883-2893.

doi: 10.2527/jas.2011-4900 originally published online July 10, 2012

The online version of this article, along with updated information and services, is located on
the World Wide Web at:

<http://www.journalofanimalscience.org/content/90/9/2883>



American Society of Animal Science

Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross¹

Y. Ramayo-Caldas,^{*2} A. Mercadé,[†] A. Castelló,^{*} B. Yang,^{*3} C. Rodríguez,[‡] E. Alves,[‡] I. Díaz,[#] N. Ibáñez-Escriche,[§] J.L. Noguera,[§] M. Pérez-Enciso,^{*†} || A.I. Fernández,[‡] and J.M. Folch^{*†}

^{*}Centre for Research in Agricultural Genomics (CRAG), Consortium CSIC-IRTA-UAB-UB. Edifici CRAG, Campus Universitat Autònoma Barcelona. 08193 Bellaterra, Spain; [†]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona. 08193 Bellaterra, Spain; [‡]Departamento de Mejora Genética Animal, INIA, Ctra. De la Coruña km. 7. 28040 Madrid, Spain; [§]IRTA, Genètica i Millora Animal. 25198 Lleida, Spain; [#]IRTA, Tecnologia dels Aliments. 17121 Monells, Spain; and ||ICREA, Passeig Lluís Companys. Barcelona, Spain.

ABSTRACT: The lipid content and fatty acid (FA) profile have an important impact in human health as well as in the technological transformation and nutritional and organoleptic quality of meat. A genome-wide association study (GWAS) on 144 backcross pigs (25% Iberian × 75% Landrace) was performed for 32 traits associated with intramuscular FA composition and indices of FA metabolism. The GWAS was carried out using Qxpak 5.0 and the genotyping information obtained from the Porcine SNP60K BeadChip (Illumina Inc., San Diego, CA). Signals of significant association considering a false-discovery rate (q -value < 0.05) were observed in 15 of the 32 analyzed traits, and a total of 813 trait-associated

SNP (TAS), distributed in 43 chromosomal intervals on almost all autosomes, were annotated. According to the clustering analysis based on functional classification, several of the annotated genes are related to FA composition and lipid metabolism. Some interesting positional concordances among TAS and previously reported QTL for FA compositions and/or other lipid traits were also found. These common genomic regions for different traits suggest pleiotropic effects for FA composition and were found primarily on SSC4, SSC8, and SSC16. These results contribute to our understanding of the complex genetic basis of FA composition and FA metabolism.

Key words: candidate gene, fatty acid, genome-wide association study, intramuscular fat, meat quality

© 2012 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2012.90:2883–2893
doi:10.2527/jas2011-4900

INTRODUCTION

¹This work was funded by MICINN Project AGL2008-04818-C03/GAN (Spanish Ministry of Science) and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036, Centre for Research in Agrigenomics). We are indebted to Nova Genetica (Lleida, Spain) and its staff for their technical support. Y. Ramayo-Caldas was funded by a FPU Ph.D. grant from the Spanish Ministerio de Educación (AP2008-01450). We thank Joaquin Casellas (UAB) and Jordi Estelle (INRA Jouy-en-Josas) for their comments, and also Martien Groenen (Wageningen, NL) for providing information about SNP positioning in Assembly 10. This manuscript has been proofread by Chuck Simons, a native English-speaking university instructor of English.

² Corresponding author: yulixis.ramayo@uab.es

³Current address: Key Laboratory for Animal Biotechnology of Jiangxi Province and the Ministry of Agriculture of China, Jiangxi Agricultural University, Nanchang, China.

Received November 9, 2011.

Accepted March 2, 2012.

Pork production is an important source of human food and accounts for more than 40% of the meat produced worldwide (Rothschild and Ruvinsky, 2011). The lipid content and fatty acid (FA) profile have an important impact in the technological transformation and the nutritional and organoleptic quality of meat. In pigs, differences in FA composition of subcutaneous fat have an important effect on fat quality as defined in terms of firmness and the degree of cohesiveness between lean and fat tissues. In muscle, FA composition has a role in the tenderness and juiciness of cooked meat, affecting its oxidative stability during processing (Wood et al., 2008). The environmental and genetic

effects on FA composition have been studied in pigs, showing moderate to high heritabilities in different tissues (Casellas et al., 2010; Ntawubizi et al., 2010).

Domestic animal populations have unique features that facilitate the genetic dissection of complex traits. Classical, family-based linkage analysis with microsatellites has been used for the detection of QTL. However, recently, new advances in high-throughput methods have allowed for the genotyping of a large number of SNP throughout the genome. Genome-wide association studies (**GWAS**) using dense marker maps are able to exploit linkage disequilibrium (**LD**) to map QTL. In comparison with the classical QTL mapping, GWAS has the advantage of using all recombination events after the mutations occurred and, therefore, increases the precision of the QTL position estimates and reduces their confidence intervals (Meuwissen and Goddard, 2000). Moreover, it considers the contribution of the variability within breeds or lines, whereas linkage analysis usually ignores it and, as a consequence, its power is decreased (Ledur et al., 2010).

Although several studies have identified QTL for FA composition and FA metabolism in pigs (Clöp et al., 2003; Nii et al., 2006; Sanchez et al., 2007; Guo et al., 2009; Uemoto et al., 2009; Quintanilla et al., 2011), a GWAS analysis using genotypic information from the Porcine SNP60K BeadChip (Illumina Inc., San Diego, CA) has not been reported.

The goal of this study is to perform a GWAS to identify chromosomal regions and positional candidate genes associated with the profile of intramuscular FA composition and indices of FA metabolism in an Iberian × Landrace cross.

MATERIALS AND METHODS

As mentioned below, the experiments were performed in Europe following national and institutional guidelines for the ethical use and treatment of animals in experiments. In addition it was approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries)

Animal Material and Analyzed Traits

The population studied descends from the crossing of three Iberian (Guadyerbas line) boars with 31 Landrace sows (**IBMAP**) from the experimental farm of Nova Genètica S. A. in Lleida, Spain (Pérez-Enciso et al., 2000; Clöp et al., 2003). Seventy-nine F1 individuals (6 males and 73 females) were obtained, and 5 of the 6 F1 males were backcrossed with 26 Landrace sows (**BC1_LD**). Here, we report results based on 144 BC1_LD (25% Iberian × 75% Landrace) pigs from 26

full-sib families. All pigs were raised and fed under the standard, intensive system in Europe; males were not castrated. After a suckling period of between 23 and 28 d, piglets were allocated in pens with 12 individuals in each pen and were given ad libitum access to a pelleted diet (13.4 MJ/kg of ME, 18.3% of CP, 1.2% of lysine). When the piglets were about 75 d old, they were moved to a fattening building. They were penned in groups of 10 to 12 animals separated by sex, and during the whole test period they had ad libitum access to a cereal-based commercial diet (13.4 MJ/kg of ME, 17.5% CP, 1% lysine). Pigs tested at the same time and in the same fattening building were considered as 1 contemporary group (batch). They were slaughtered according to national and institutional guidelines for the ethical use and treatment of animals in experiments at an average age of 179.8 ± 2.6 d. For all individuals, a 250 g sample of the LM was collected from the left half-carcass in the slaughter line, with 50 g of it quick-frozen in liquid nitrogen until genomic laboratory processing.

Meat analyses were performed on 200-g samples of the longissimus dorsi muscle at the IRTA-Centre of Food Technology (Lleida, Spain). The intramuscular percentage of FA (**IMFA**) was measured by Near Infrared Transmittance (**NIT**; Infratec 1625, Tecator, Hoganas, Sweden). A protocol based on gas chromatography of methyl esters (Mach et al., 2006) was employed to determine FA composition in the C:12 to C:22 range; subsequently, the percentage of each individual FA content was calculated, along with the global percentage of SFA, MUFA, and PUFA.

In total, 32 traits were analyzed: 15 for IM FA composition and 17 for indices of FA metabolism, including ratios of FA as indices for desaturase and elongase enzymes activities (Table 1).

Genotyping and Quality Control

A total of 197 animals, including the founder populations, were genotyped for 62,163 SNP with the Porcine SNP60K BeadChip (Ramos et al., 2009) and using the Infinium HD Assay Ultra protocol (Illumina Inc., San Diego, CA) and were visualized with the GenomeStudio software (Illumina). PLINK (Purcell et al., 2007) software was used to remove markers that showed minor allele frequency (**MAF**) < 5% and missing genotypes > 5%. The Pedstats program (Wigginton and Abecasis, 2005) was employed to check Mendelian inheritance errors. The SNP located in sex chromosomes and those not mapped in the Sscrofa10 assembly were also excluded. Therefore, a subset of 48,119 SNP was selected for subsequent data analysis.

Table 1. Descriptive statistics including mean, SD and estimated heritability (h^2) of fatty acid (FA) composition and FA indices

Item	Trait	Label	Mean	SD	h^2
SFA	C14:0	Myristic acid	1.18	0.15	0.54
	C16:0	Palmitic acid	22.60	1.20	0.57
	C17:0	Heptadecanoic acid	0.27	0.07	0.30
	C18:0	Stearic acid	14.18	1.03	0.24
	C20:0	Arachidic acid	0.25	0.08	0.42
MUFA	C16:1(n-7)	Palmitoleic acid	2.50	0.39	0.55
	C17:1	Heptadecenoic acid	0.27	0.11	0.41
	C18:1(n-9)	Oleic acid	40.08	2.76	0.40
	C18:1(n-7)	Octadecenoic acid	3.88	0.36	0.42
	C20:1(n-9)	Eicosenoic acid	0.85	0.11	0.47
PUFA	C18:2(n-6)	Linoleic acid	10.35	2.37	0.22
	C18:3(n-3)	α -Linolenic acid	0.65	0.29	0.18
	C20:2(n-6)	Eicosadienoic acid	0.54	0.12	0.26
	C20:3(n-6)	Eicosatrienoic acid	0.28	0.13	0.16
	C20:4(n-6)	Arachidonic acid	1.54	0.73	0.40
Metabolic ratios	ACL	Average Chain Length	17.47	0.05	0.41
	Σ SFA	SFA	38.47	1.63	0.53
	Σ MUFA	MUFA	47.96	3.06	0.19
	Σ PUFA	PUFA	13.36	3.29	0.18
	PI	Peroxidability indices	20.00	5.72	0.40
	DBI	Double-bond indices	0.78	0.09	0.41
	UI	Unsaturated indices	2.04	0.27	0.19
	PUFA/SFA	Ratio of PUFA to SFA	0.35	0.09	0.18
	MUFA/SFA	Ratio of MUFA to SFA	1.25	0.10	0.35
	FA ratios				
$\Delta 9$ -desaturase activity	C16:1(n-7)/C16:0		0.11	0.01	0.51
	C18:1(n-9)/C18:0		2.84	0.29	0.35
	C20:1/C20:0		3.64	1.06	0.42
$\Delta 5$ -desaturase activity	C20:4(n-6)/C20:3(n-6)		5.86	3.02	0.37
Elongase activity	C18:1(n-7)/C16:1(n-7)		1.58	0.19	0.57
	C20:2(n-6)/C18:2(n-6)		0.05	0.01	0.39
Combined effect of desaturase and elongase	C20:3(n-6)/C18:2(n-6)		0.03	0.01	0.15
	C20:4/C18:2(n-6)		0.14	0.04	0.39

Genome-Wide Association Analysis

A GWAS on IM FA composition and indices of FA metabolism was performed in 144 BC1_LD animals. A mixed model (Henderson, 1984, 1975) accounting for additive effects (see below) was performed using Qxpak 5.0 (Pérez-Enciso and Misztal, 2011) using the formula:

$$y_{ijklkm} = \text{Sex}_i + \text{Batch}_j + \beta c_l + \lambda_l a_k + u_l + e_{ijklkm},$$

in which y_{ijklkm} is the l th individual record, sex (2 categories) and batch (5 categories) are fixed effects, β is a covariate coefficient with c being carcass weight, λ_l is a $-1, 0, +1$ indicator variable depending on the l th individual genotype for the k th SNP, a_k represents the additive effect associated with the SNP, u_l represents the infinitesimal genetic effect treated as random and distributed as $N(0, A\sigma_u)$ where A is a numerator of

kinship matrix, and e_{ijklkm} is the residual. The R package q-value (Storey and Tibshirani, 2003) was used to calculate the false-discovery rate (FDR), and the cut-off of the significant association at the whole-genome level was set at $q\text{-value} \leq 0.05$.

To obtain, approximately, the variance explained by each SNP, we first fitted a model that included all associated SNP for each trait. This is important because, otherwise, SNP effect estimates are inflated by confounding effects (e.g., LD among SNP). The fraction of the phenotypic variance explained by SNP $_t$ ψ_t was calculated as:

$$\psi_t = \frac{2p_t(1-p_t)\hat{a}_t^2}{\hat{\sigma}_e^2 + \hat{\sigma}_a^2 + \sum_{n=1}^n 2p_n(1-p_n)\hat{a}_n^2}$$

in which p_t is the allele frequency of t th SNP in the population studied, with an estimated effect \hat{a}_t and assuming there are n significant SNP associated with the

trait; and σ_e^2 and σ_a^2 are the estimated residual and the additive infinitesimal variances, respectively.

Version 2.12.1 of R (Ihaka and Gentleman, 1996; <http://cran.r-project.org>) was used to calculate the descriptive statistics for the 32 analyzed traits and the phenotypic correlations for traits that showed significantly associated SNP. Oleic FA was also included in the correlation analysis due to its importance in meat FA composition. Before analyzing the correlation, data were adjusted for fixed effect (sex, batch) and carcass weight as a covariate with the linear models procedure of R.

Gene Annotation and Functional Classification

The significantly associated SNP ($q\text{-value} \times 0.05$) were initially mapped in the Sscrofa10 assembly. Then, the putative candidate chromosomal regions were defined by finding those genome locations comprising windows of at least 2 consecutive and significant SNP. The SNP positioned within the candidate chromosomal regions were also determined in Sscrofa9 assembly (Pig Ensembl Genome Browser, http://www.ensembl.org/Sus_scrofa/Info/index). Gene annotations were retrieved from the Ensembl Genes 62 Database using Biomart software (<http://www.biomart.org>; Flicek et al., 2010), assuming 1 Mb downstream/upstream around the candidate chromosomal regions. The functional classification of genes was performed using the DAVID online annotation database (<http://david.abcc.ncifcrf.gov>; Huang et al., 2008, 2009). Those genes functionally related to lipid and FA metabolism were retained as positional candidate genes. The genomic coordinates of the annotated genes in the Ensembl Sscrofa10 dataset were validated by BLAST analysis (www.animalgenome.org).

RESULTS AND DISCUSSION

Phenotype Statistics and GWAS Results

The descriptive statistics of IMF FA composition and indices of FA metabolism are reported in Table 1. In accordance with previous reports (Yang et al., 2010; Quintanilla et al., 2011), moderate positive phenotype correlations were observed between myristic, palmitic and palmitoleic FA ($r_{C14:0,C16:0} = 0.68$, $r_{C14:0,C16:1} = 0.61$, $r_{C16:0,C16:1(n-7)} = 0.58$). As expected, oleic acid showed a strong positive correlation with MUFA ($r = 0.98$) and a negative correlation with PUFA ($r = -0.87$). Conversely, oleic acid showed a clearly negative correlation with linoleic acid ($r = -0.85$), which also has been reported by Quintanilla et al. (2011) in gluteus medius and longissimus thoracis et lumborum muscles. The relationship between enzyme activity indices showed a moderate positive correlation for Δ^9

–desaturase activity ($r_{C16:1(n-7)/C16:0, C18:1(n-9)/C18:0} = 0.70$; Table 1), as previously reported in the LM of commercial pigs (Ntawubizi et al., 2010). In agreement with previous reports (Colman et al., 2008; Casellas et al., 2010; Ntawubizi et al., 2010; Sellier et al., 2010), the estimated heritabilities (h^2) were moderately large, ranging from 0.15 to 0.57 (Table 1).

A total of 144 BC1_LD animals passed quality control, and a subset of 48,119 SNP with high-genotyping quality (call rate > 0.99) was selected for GWAS. We were able to detect significant association signals ($q\text{-value} < 0.05$) in 15 out of the 32 traits analyzed. A total of 813 significant trait-associated SNP (TAS), located in 43 chromosomal regions distributed among almost all autosomes, were annotated according to the Sscrofa9 assembly (Table 2). Of the 813 TAS, 596 (73.3%) were mapped in intergenic regions, 189 (23.3%) were within genes, and 28 (3.4%) were unmapped in the Sscrofa9 assembly. Most of the gene-associated SNP were located in introns (93.6%): 7 were exonic and synonymous; and 5 were exonic, nonsynonymous SNPs. Annotation of the 5 genes containing nonsynonymous polymorphisms showed that 2 correspond to uncharacterized proteins (ENSSSCG00000004415 and ENSSSCG00000006022) and the rest to genes that were not directly related to the FA metabolism: Metadherin (MTDH, ENSSSCG00000006086), Serine incorporator 2 (SERINC2, ENSSSCG000000035959) and RAD17 homolog (*S. pombe*; RAD17, ENSSSCG00000016960). A more detailed description is presented in Table 2, including the 2 flanking genes around the most significant SNP for each trait.

A clustering analysis of the genes located in TAS regions revealed different functional categories related to lipid metabolism (Supplemental Table 3; <http://journalofanimalscience.org>). These positional candidate genes are discussed below.

Genomic regions associated with the composition of FA in LM were found for some SFA, MUFA, and PUFA. Significant association was found for the SFA, myristic (on SSC5 and SSC17), palmitic (on 8 different chromosomes, Table 2), and arachidic acids (SSC16). Within MUFA, 6 genomic regions (SSC4, SSC5, SSC6, SSC7, SSC8, and SSC15) were associated with palmitoleic acid amounts and 4 (SSC1, SSC4, SSC6, and SSC11) with oleic (Table 2). For PUFA, only linoleic acid composition was significant. However, this was the trait showing the largest number of associated positions along the pig genome, with significant association signals in 11 out of the 18 autosomes. Mammals lack the $\Delta12$ and $\Delta15$ -desaturases required to synthesize FA of the n-6 and n-3 series. As a consequence, the polyunsaturated linoleic and α -linoleic acids must be provided by diet, and their tissue concentrations respond

Table 2. Description of the 43 putative candidate chromosomal intervals

Interval	Chromosome	Position Mb ¹ /Start-End	N _{SNP}	Most Significant SNP ³	P-value	q-value	Ib ⁴	Ld ⁵	ψ ⁷	Associated Traits
1	1	31.9-32.1	8	ALGA0002268	6.0x10 ⁻⁶	2.5x10 ⁻²	1	0.21	0.009	*SFA, UI
2	1	44.8-49.4	23	ASGA0002580	3.4x10 ⁻⁵	2.0x10 ⁻²	1	0.11	0.002	*C16:0, C18:2(n-6), SFA, UI
3	1	55.3-59.2	57	ASGA0003001	1.9x10 ⁻⁵	3.0x10 ⁻²	1	0.44	0.026	C16:0, SFA, *UI
4	1	72.7-78.3	97	H3GA0001996	7.7x10 ⁻⁵	4.0x10 ⁻²	1	0.15	0.019	*SFA, C16:0
5	1	126.6-130.8	46	DRGA0001507	7.6x10 ⁻⁵	3.0x10 ⁻²	0.83	0.32	0.026	*C18:2(n-6), MUFA/SFA
6	1	137.2-140.2	34	ALGA0006048	6.9x10 ⁻⁵	4.0x10 ⁻²	0.67	0.38	0.025	*C18:1(n-9), C18:2(n-6), PUFA, PUFA/SFA
7	2	19.1-23.8	6	DRGA0002840	1.3x10 ⁻⁶	6.1x10 ⁻³	1	0.06	0.010	C16:0
8	3	90.1-94.6	4	ASGA0015593	1.4x10 ⁻⁵	3.6x10 ⁻²	1	0.06	0.003	C18:1(n-7)/C16:1(n-7)
9	3	108.8-111.8	2	ASGA0101714	5.3x10 ⁻⁵	4.1x10 ⁻²	1	0.49	0.033	C16:1(n-7)/C16:0
10	4	10.9-14.5	24	ASGA0018461	8.0x10 ⁻⁵	3.0x10 ⁻²	0.67	0.23	0.031	*MUFA, C18:2(n-6), C18:1(n-9)
11	4	40.1-44.4	37	H3GA0012592	3.0x10 ⁻⁶	2.3x10 ⁻²	1	0.02	0.011	*PUFA, MUFA, PUFA/SFA
12	4	51.1-56.7	15	ASGA0019639	2.8x10 ⁻⁵	2.9x10 ⁻²	1	0.26	0.021	*C16:1(n-7), C18:2(n-6), C18:1(n-9)
13	4	78.2-80.2	13	ALGA0026066	4.6x10 ⁻⁵	4.6x10 ⁻²	1	0.13	0.004	MUFA/SFA
14	4	90.7-92.2	6	H3GA0013315	5.7x10 ⁻⁶	1.0x10 ⁻²	1	0.38	0.012	*C16:1(n-7), C18:2(n-6), C16:0, C18:1(n-7)/C16:1(n-7)
15	4	100.3-103.5	13	ALGA0027207	6.4x10 ⁻⁵	2.4x10 ⁻²	1	0.43	0.015	*C18:1(n-9), MUFA/SFA, MUFA
16	4	112.4-112.7	6	ASGA0021882	5.4x10 ⁻⁵	4.1x10 ⁻²	1	0.27	0.008	SFA
17	4	123.2-128.9	35	ALGA0031433	8.7x10 ⁻⁷	2.2x10 ⁻³	1	0.36	0.031	*C16:1(n-7), C16:0, C16:1(n-7)/C16:0, PUFA, PUFA/SFA
18	5	24.9-27.9	30	ALGA0031433	1.1x10 ⁻⁵	4.0x10 ⁻²	1	0.41	0.021	C14:0
19	5	70-72.1	2	ALGA0032946	2.4x10 ⁻⁵	2.7x10 ⁻²	1	0.38	0.019	C16:1(n-7)
20	6	55.6-56.9	18	ASGA0104276	1.2x10 ⁻⁶	1.3x10 ⁻²	1	0.3	0.018	*C18:1(n-9), MUFA, MUFA/SFA, C16:1(n-7), C18:2(n-6)
21	7	98.7-101.1	20	ASGA0035273	7.9x10 ⁻⁵	4.9x10 ⁻²	1	0.03	0.001	*C16:1(n-7), C20:1/C20:0
22	7	122.2-125.9	40	H3GA0023298	1.6x10 ⁻⁴	4.1x10 ⁻²	0.67	0.28	0.023	C18:2(n-6)
23	8	10.6-13.7	4	DRGA0008319	7.2x10 ⁻⁷	1.2x10 ⁻²	1	0.42	0.037	*SFA, UI
24	8	68.6-71.9	9	DRGA0008674	7.8x10 ⁻⁵	2.9x10 ⁻²	1	0.4	0.007	*C16:0, SFA
25	8	77.6-80.3	3	ALGA0048684	3.5x10 ⁻⁷	1.2x10 ⁻³	1	0.36	0.011	*C16:1(n-7), C18:1(n-7)/C16:1(n-7)
26	8	92.1-96.7	52	SIRI0000509	3.7x10 ⁻⁹	4.9x10 ⁻⁵	1	0.34	0.061	*C16:1(n-7), C16:0, C18:1(n-7)/C16:1(n-7), C16:1(n-7)/C16:0, SFA, UI
27	8	103.8-107.5	10	ASGA0039809	1.2x10 ⁻⁵	1.1x10 ⁻²	1	0.4	0.025	C16:0
28	9	18.9-21.7	11	DRGA0009211	4.2x10 ⁻⁵	1.3x10 ⁻¹	1	0.35	0.037	C20:1/C20:0
29	11	6.2-8.8	22	DRGA0010773	3.9x10 ⁻⁶	1.5x10 ⁻²	1	0.04	0.033	C18:1(n-9), C18:2(n-6), MUFA/SFA, *MUFA
30	13	78.4-83.3	10	DIAS0004396	1.3x10 ⁻⁶	1.4x10 ⁻²	1	0.47	0.044	PUFA/SFA, C16:0, C18:1(n-9), SFA, *UI
31	14	2.1-4.1	19	H3GA0038495	1.9x10 ⁻⁵	2.5x10 ⁻²	1	0.38	0.032	*C18:2(n-6), PUFA, PUFA/SFA, UI
32	14	28.4-37	21	ASGA0062744	1.8x10 ⁻⁵	3.8x10 ⁻²	1	0.3	0.008	SFA
33	15	Unmapped ⁶	2	MARCO042106	1.4x10 ⁻⁵	4.4x10 ⁻²	1	0.22	0.027	C16:1(n-7)/C16:0
34	15	24.4-28	8	ALGA0111611	4.1x10 ⁻⁶	1.0x10 ⁻²	1	0.5	0.021	C16:0
35	15	33.4-34.5	13	ALGA0110309	5.5x10 ⁻⁵	4.2x10 ⁻²	1	0.12	0.003	*C16:1(n-7), MUFA/SFA
36	15	50-50.4	2	ASGA0069654	2.6x10 ⁻⁴	4.5x10 ⁻²	1	0.18	0.005	C16:0
37	15	104.2-105.6	9	ASGA0070398	1.5x10 ⁻⁵	1.1x10 ⁻²	0.5	0.28	0.004	C16:0
38	16	39.6-45.2	15	DRGA0016162	1.5x10 ⁻⁶	1.6x10 ⁻²	1	0.46	0.013	*C20:0, C20:1/C20:0
39	16	64.3-67.9	36	ASGA0085192	1.5x10 ⁻⁴	4.1x10 ⁻²	1	0.34	0.008	C18:2(n-6)
40	17	20.1-21.9	12	H3GA0048059	8.0x10 ⁻⁵	2.9x10 ⁻²	1	0.43	0.003	*C16:0, C20:1/C20:0, C14:0
41	17	34.5-34.7	4	ALGA0094522	1.2x10 ⁻⁶	2.3x10 ⁻²	0.83	0.38	0.040	C14:0
42	17	45.3-49.5	11	ALGA0095453	1.6x10 ⁻⁶	2.3x10 ⁻²	1	0.49	0.020	*C14:0, C16:1(n-7), C18:1(n-7)/C16:1(n-7)
43	18	10.2-15.8	4	ASGA0078979	1.7x10 ⁻⁴	3.9x10 ⁻²	1	0.26	0.025	C16:0

¹Indicates the most significant trait for each chromosomal interval. The P- and q-values refer to this trait. All of the traits shown in the table are significant (q-value \leq 0.05).

²The genomic coordinates are expressed in Mb and are relative to the Sscrofa April 2009 genome sequence assembly (Sscrofa9).

³Number of significant SNP in the interval.

⁴Ib = allelic frequency in the founder Iberian boars.

⁵Ld = allelic frequency in the founder Landrace sows.

⁶The SNP that defined the interval were mapped in the Sscrofa10 assembly, but not in the Sscrofa9 (in which gene annotation is available).

⁷ ψ = proportion of the phenotypic variance explained by the most significant SNP.

rapidly to dietary changes. Therefore, loci associated with linoleic acid content may have an effect on its absorption or transformation into longer-chain PUFA rather than on its biosynthesis.

Concordance of TAS with QTL

It was investigated herein whether the chromosomal regions delimited by the significant SNP described in the present study overlap with QTL previously reported in the IBMAP cross (Pérez-Enciso et al., 2000; Clop et al., 2003). In addition, the concordance between TAS and QTL for the profile of FA or other lipid traits deposited in the pig QTL database (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>) was analyzed. Twenty-three of the 43 chromosomal intervals detected in this study (53.5%) showed positional concordance with QTL for FA composition and/or other lipid traits reported in the pig QTL database (Supplemental Table 4; <http://journalofanimalscience.org>). The chromosomal regions in which concordant QTL have been reported for fatty acid metabolism are discussed in detail below; a full description is in Supplementary Table 4 (<http://journalofanimalscience.org>).

SSC1

The 44.8–49.4 Mb region of SSC1 was associated with the percentages of palmitic acid, linoleic acid, SFA, and unsaturated indices (UI). Single nucleotide polymorphism ASGA0002580 showed the greatest significance for palmitic acid (P value = 3.4×10^{-5} ; Table 2). This region overlaps with QTL reported for total lipid content in LM (Malek et al., 2001; Pig QTLdb Id: 79) and low density lipoprotein (LDL) cholesterol concentrations (Gallardo et al., 2008; Pig QTLdb Id: 7709).

SSC4

In SSC4, the 51.1–56.7 Mb region was associated with percentages of palmitoleic, oleic, and linoleic acids (Table 2), SNP ASGA0019639 being the most significantly associated for palmitoleic acid (P value = 2.8×10^{-5}), which explained 2.1% of the phenotypic variance of this trait (Table 2). In previous works, an F2-cross between Iberian (Guadyerbas) boars and Landrace sows was analyzed and QTL for backfat thickness (Varona et al., 2002), the percentage of oleic and linoleic acids, the double-bond index (DBI) and the peroxidability index (PI; Pérez-Enciso et al., 2000; Clop et al., 2003) were reported. Several positional candidate genes were analyzed for this QTL, including 2,4-dienoyl-CoA reductase (DECRI; Clop et al., 2003;

Amills et al., 2005), FA binding protein 4 (FABP4; Mercadé et al., 2006), and FABP5 (Estellé et al., 2006; Ojeda et al., 2008). However, these studies indicated that the SNP analyzed in candidate genes are not the causative mutation of the detected QTL. An association analysis with FABP4–FABP5 haplotypes suggested that the QTL for backfat thickness is likely located between these 2 genes. Here, we have analyzed the IMF FA composition in BC1_LD animals (a backcross of 25% Iberian and 75% Landrace). Remarkably, a SNP (H3GA0012720; P value = 5.2×10^{-5}) within the significant chromosomal interval mapped at intron 1 of the FABP5 gene was only at 1.8 kb of distance from the SNP FABP5:g.3000T>G reported by Estellé et al. (2006). The observed concordance suggests a pleiotropic effect of this locus on both fat deposition and FA composition, which may affect both intramuscular fat and backfat thickness.

A different SNP-defined region (90.7–92.2 Mb) on SSC4 was associated with palmitic, palmitoleic, and linoleic acids and the C18:1(n-7)/C16:1(n-7) ratio. Many QTL have been reported in this region of SSC4. In the IBMAP cross, Pérez-Enciso et al. (2000) identified QTL for palmitic and linoleic acids and UI (Pig QTLdb Id: 530, 535, 541); Grindflek et al. (2001) for heptadecanoic acid; and Guo et al., (2009) for linoleic, MUFA, PUFA, and linolenic acids (Pig QTLdb Ids: 6392, 6393, 6394, 6395). The most significant SNP within this interval was H3GA0013315 (P value = 5.7×10^{-6}), which is located at ≈ 1.1 Mb of the apolipoprotein A-II gene (APOA2) and explained 1.2% of the phenotypic variance for this trait. This gene encodes the second most abundant protein of the high-density lipoprotein (HDL) particles and is one of the target genes of the PPAR signaling pathway (Blanco-Vaca et al., 2001). In addition, a polymorphism in the promoter region of the gene interacts with the saturated fat intake to influence body mass index in humans (Corella et al., 2009). The APOA2 gene has been linked to the plasma concentrations of the APOA2 protein and FFA in mice and humans (Warden et al., 1993). Upstream transcription factor 1 (USF1) is located within the same chromosomal interval. The protein encoded by this gene is one of the major components of the complexes that bind to the insulin response sequence of the FA synthase (FAS) promoter. Hence, it may play an important role in the regulation of FAS (Wong and Sul, 2010).

SSC8

A strong association signal was found in SSC8 (92.1–96.7 Mb) with palmitic acid, palmitoleic acid, SFA, and C16:1(n-7)/C16:0 and C18:1(n-7)/C16:1(n-7) ratios (Figure 1, Table 2). These results suggest the presence of

genes that regulate both the desaturation and elongation processes of FA. In this chromosomal region, SNP SIRI0000509 was the most significant for palmitic acid (P -value = 8.2×10^{-8}), palmitoleic acid (P -value = 3.7×10^{-9}) and the C18:1(n-7)/C16:1(n-7) ratio (P -value = 1.8×10^{-8}). Conversely, the most significant SNP for the C16:1(n-7)/C16:0 ratio was ALGA0049269 (P -value = 8.3×10^{-7}), located at 1.8 Mb of SIRI0000509, according to Sscrofa9 assembly. It should be noted that SNP SIRI0000509 showed the greatest contribution to the phenotypic variance, explaining 6.1% of the palmitoleic acid variation (Table 2). A positional concordance was observed between this region and QTL for palmitic (Pig QTLdb Id: 469) and palmitoleic (Pig QTLdb Id: 470) acids reported in backfat (Clou et al., 2003). Recently, Yang et al. (2011), using a casual phenotype network, suggested that this QTL has a direct effect on palmitic and palmitoleic acids but only an indirect association with oleic acid in both the IBMAP and Duroc \times Erhualian crosses. It is worth noting that a nonsynonymous polymorphism in the microsomal triglyceride transfer protein gene (**MTTP**) was associated with palmitic, palmitoleic, and oleic acids and also to the lipid transfer activity of the MTTP protein (Estellé et al., 2009). This gene is located within the QTL region and is involved in the transport of triglycerides into endoplasmic reticulum to form lipid droplets. However, it should be noted that 3 additional candidate genes are located proximal to the markers mentioned. The cytochrome P450 family 2 subfamily U polypeptide 1 gene (**CYP2U1**) is located \approx 334 kb from SNP ALGA0049269 and \approx 2.1 Mb from SNP SIRI0000509. This gene is involved in the arachidonic acid signaling pathway and catalyzes the hydroxylation of arachidonic, docosahexaenoic, and other long-chain FA (Chuang et al., 2004). In addition, the phospholipase A2 group XIA (**PLA2G12A**), a gene involved in the same pathway, is located \approx 863 kb from SNP ALGA0049269 and \approx 935 kb from SNP SIRI0000509; PLA2G12A is a member of the phospholipase A2 (**PLA2**) family, which catalyzes the hydrolysis of the sn-2 position of membrane glycerophospholipids to liberate FFA and lysophospholipids (Murakami et al., 2010). Finally, the hydroxyacyl-CoA dehydrogenase gene (**HADH**) was located \approx 278 kb from SNP ALGA0049269 and \approx 2.1 Mb from SNP SIRI0000509. This gene is involved in the mitochondrial FA beta-oxidation pathway and has an elevated activity with medium-chain-length FA (Safran et al., 2010).

SSC14

The 28.4–37 Mb genomic region on SSC14 was associated with the SFA content, ASGA0062744 being the SNP showing the strongest association signal

(P -value = 1.8×10^{-5}) and explaining 0.8% of the phenotypic variance (Table 2). In this position, QTL for PUFA (Pig QTLdb Id: 7446) and the PUFA:SFA ratio (Pig QTLdb Id: 7447) have previously been reported in LM (Sanchez et al., 2007). One of the most interesting candidate genes within this interval is the acetoacetyl-CoA synthetase gene (**AACS**), a member of the acyl-CoA synthetases that catalyze the activation of FA by esterification with CoA. This initial activation reaction is fundamental for both the anabolic and catabolic pathways of FA (Watkins et al., 2007).

SSC16

In SSC16 (39.6–45.2 Mb), a strong association signal for arachidic acid and the C20:1/C20:0 ratio was found (Figure 2, Table 2). A QTL for arachidic acid has been reported in the same region, both in LM and abdominal fat (Guo et al., 2009). In addition, a QTL for myristic acid was reported by Uemoto et al. (2009) in the same interval. The most significant SNP were DRGA0016176 for arachidic acid (P -value = 1.5×10^{-6}) and ALGA0090471 for the ratio of C20:1/C20:0 (P -value = 5.5×10^{-5}). Interestingly, within this interval, \approx 221 kb from SNP ALGA0090471 and \approx 2.9 Mb from SNP DRGA0016176, the ELOVL family member 7 elongation of long-chain FA (**ELOVL7**) gene is positioned. This gene is a member of the elongation of the very-long-chain FA gene family and mainly uses SFA and MUFA as substrates (Guillou et al., 2010).

It should be noted that full concordance is not expected between the chromosomal regions identified in this study and the QTL previously reported in the IBMAP cross. Previous FA QTL were measured in the backfat of F2 cross animals, whereas here IM FA in backcross animals have been analyzed. The method of analysis used was also different; previously reported results were obtained with linkage analyses and microsatellite markers, whereas here an association method was employed. Nevertheless, linkage disequilibrium and linkage signals are partly confounded because of the material analyzed (a backcross). This is particularly true if allelic frequencies are very different among founders, and this was observed for a large amount of the most significant SNP (Table 2).

Comparison between TAS and QTL of the pig QTL database showed common genomic regions for different, although related, traits, suggesting pleiotropic effects. Conversely, other studies proposed that the genetic architecture of FA composition is mainly different between tissues (Nii et al., 2006; Guo et al., 2009; Quintanilla et al., 2011), with the exception of the QTL on SSC16 for arachidic acid in abdominal fat and LM (Guo et al., 2009) and QTL on SSC6 and SSC7

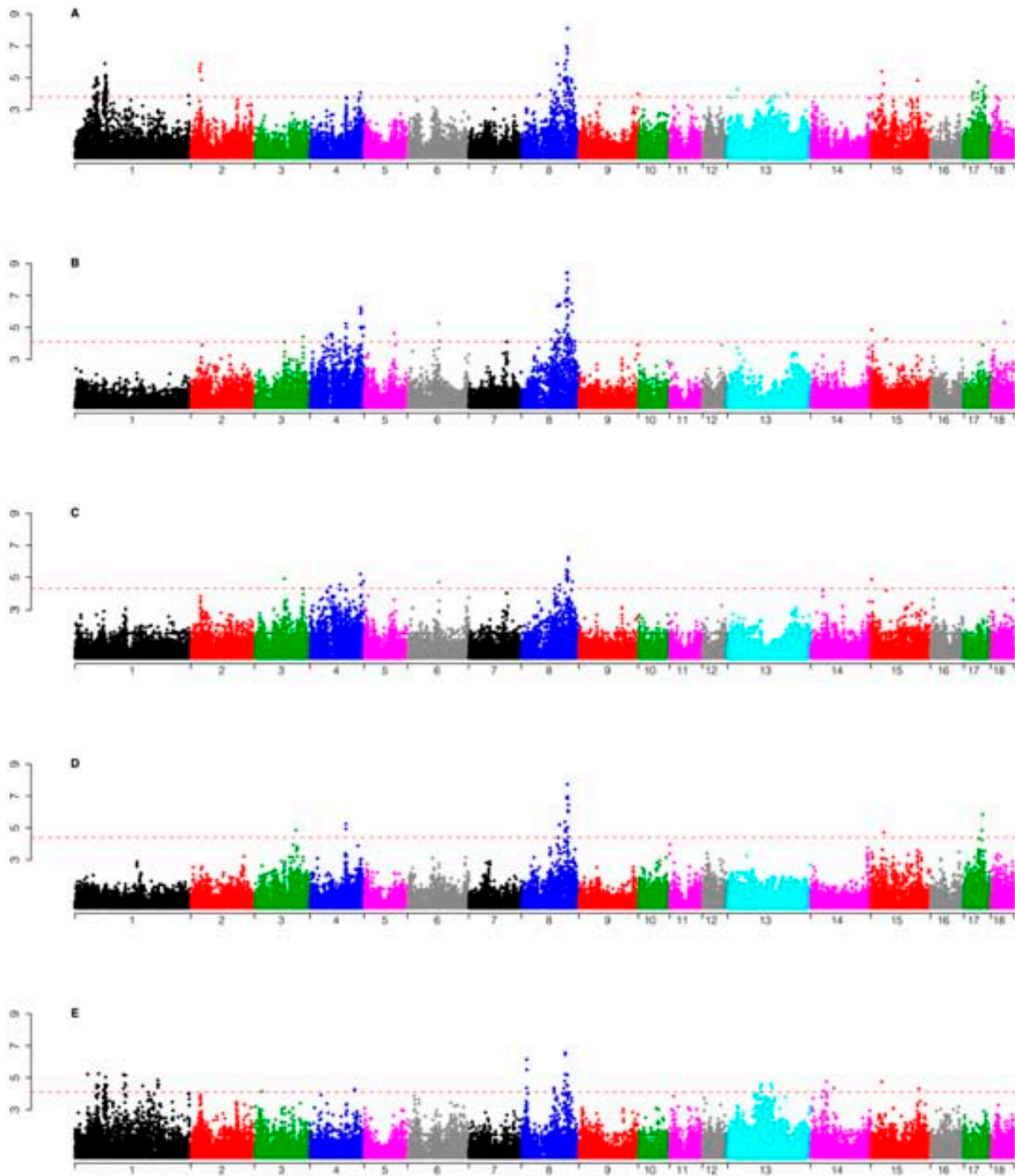


Figure 1. Manhattan plot of the genome-wide association study (GWAS) result for A) palmitic acid, B) palmitoleic acid, C) ratios of C16:1(n-7)/C16:0, D) ratios of C18:1(n-7)/C16:1(n-7), and E) SFA. The X-axis represents the chromosomes, and the Y-axis shows the $-\log_{10}(P\text{-value})$. The horizontal, dashed line represents the cut-off of the significant association at the whole-genome stage ($q\text{-value} < 0.05$). See online version for figure in color.

(Quintanilla et al., 2011) for cis-vaccenic in two different muscles (gluteus medius and longissimus thoracis).

The results obtained here will enhance our knowledge of the genetic basis of FA composition and metabolism. This information can be used in the selection of commercial pig breeds through marker-assisted selection to improve meat quality without affecting production yield. In addition, pig is an excellent

biomedical model, and the identification of genes and polymorphisms that regulate lipid metabolism in pigs will likely be important for the study of diseases such as obesity, diabetes, and atherosclerosis in humans.

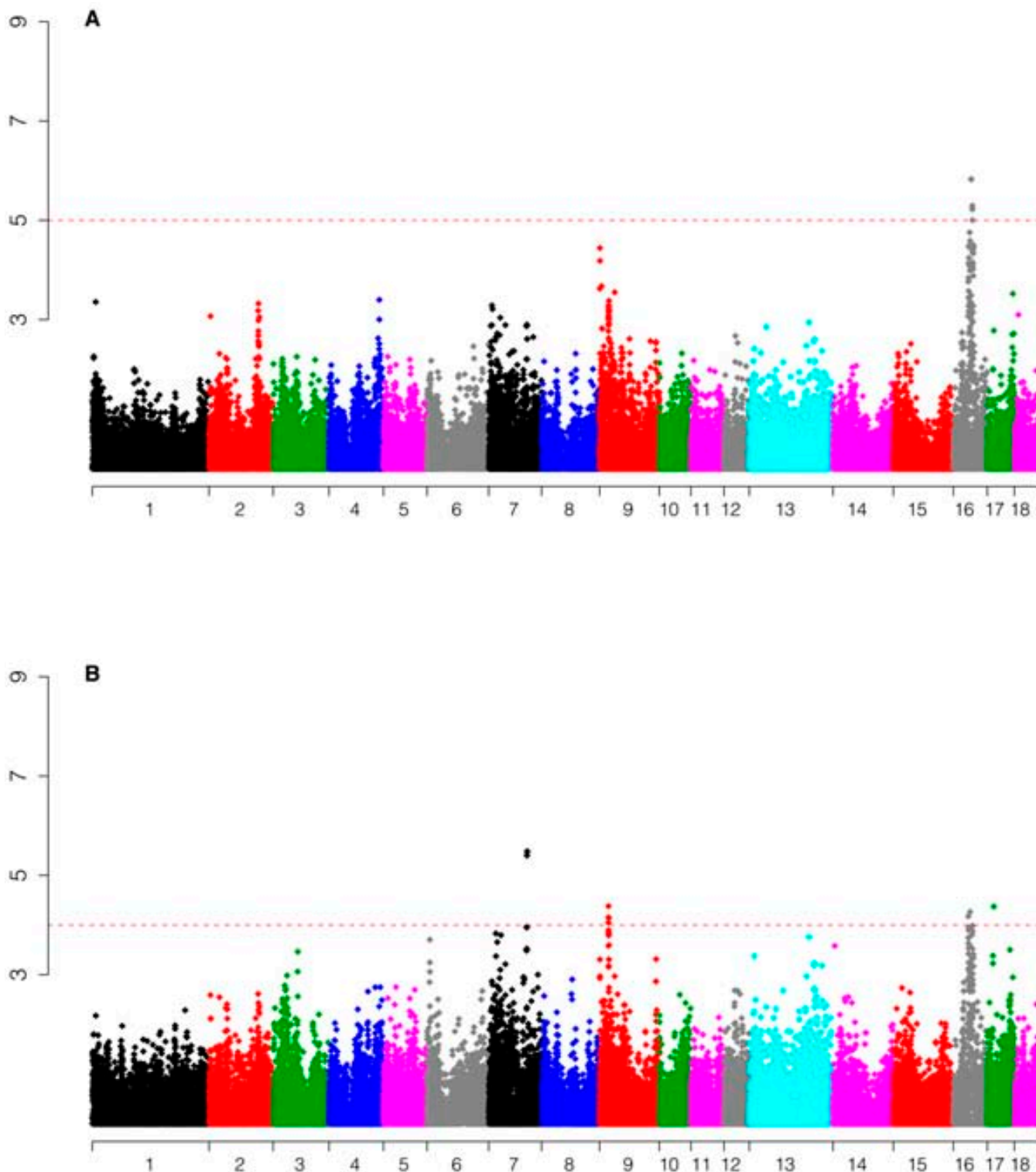


Figure 2. Manhattan plot of the genome-wide association study (GWAS) result for A) arachidic acid and B) ratios of C20:1/C20:0. The X-axis represents the chromosomes, and the Y-axis shows the $-\log_{10}(P\text{-value})$. The horizontal, dashed line represents the cut-off of the significant association at the whole-genome stage ($q\text{-value} < 0.05$). See online version for figure in color.

Conclusions

A GWAS was performed using the Porcine SNP60K BeadChip genotypes and the profile of intramuscular FA composition and indices of FA metabolism in a backcross of Iberian \times Landrace pigs. A total of 43 chromosomal

regions containing 813 TAS were identified at a $q\text{-value} < 0.05$. The annotation of these genomic regions revealed genes that are related to FA metabolism. A relevant concordance was detected between associated genomic regions and previously reported QTL for FA and different lipid traits, mainly in SSC4, SSC8 and

SSC16. These results suggest pleiotropic effects of these QTL and contribute to our understanding of the genetic basis of FA composition.

LITERATURE CITED

- Amills, M., O. Vidal, L. Varona, A. Tomás, M. Gil, A. Sánchez, and J. L. Noguera. 2005. Polymorphism of the pig 2,4-dienoyl CoA reductase 1 gene (DECR1) and its association with carcass and meat quality traits. *J. Anim. Sci.* 83:493–498.
- Blanco-Vaca, F., J. C. Escola-Gil, J. M. Martín-Campos, and J. Julve. 2001. Role of apoA-II in lipid metabolism and atherosclerosis: Advances in the study of an enigmatic protein. *J. Lipid Res.* 42:1727–1739.
- Casellas, J., J. L. Noguera, J. Reixach, I. Díaz, M. Amills, and R. Quintanilla. 2010. Bayes factor analyses of heritability for serum and muscle lipid traits in Duroc pigs. *J. Anim. Sci.* 88:2246–2254.
- Chuang, S. S., C. Helvig, M. Taimi, H. A. Ramshaw, A. H. Collop, M. Amad, J. A. White, M. Petkovich, G. Jones, and B. Korczak. 2004. CYP2U1, a novel human thymus- and brain-specific cytochrome P450, catalyzes ω - and (ω -1)-hydroxylation of fatty acids. *J. Biol. Chem.* 279:6305–6314.
- Clop, A., C. Ovilo, M. Pérez-Enciso, A. Cercos, A. Tomás, A. Fernandez, A. Coll, J. M. Folch, C. Barragan, I. Díaz, M. A. Oliver, L. Varona, L. Silió, A. Sánchez, and J. L. Noguera. 2003. Detection of QTL affecting fatty acid composition in the pig. *Mamm. Genome* 14:650–656.
- Colman, E., S. Janssens, M. Ntawubizi, K. Raes, N. Buys, and S. De Smet. 2008. Genetic parameters for intramuscular fatty acid composition and metabolism in pork. 4 p. In: 54th International Congress of Meat Science and Technology, Cape Town, South Africa.
- Corella, D., G. Peloso, D. K. Arnett, S. Demissie, L. A. Cupples, K. Tucker, C.-Q. Lai, L. D. Parnell, O. Coltell, Y.-C. Lee, and J. M. Ordovas. 2009. APOA2, dietary fat, and body mass index: Replication of a gene-diet interaction in 3 independent populations. *Arch. Intern. Med.* 169:1897–1906.
- Estellé, J., A. I. Fernández, M. Pérez-Enciso, A. Fernández, C. Rodríguez, A. Sánchez, J. L. Noguera, and J. M. Folch. 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. *Anim. Genet.* 40:813–820.
- Estellé, J., M. Pérez-Enciso, A. Mercadé, L. Varona, E. Alves, A. Sánchez, and J. M. Folch. 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.
- Flicek, P., M. R. Amode, D. Barrell, K. Beal, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. Kahari, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, P. Larsson, I. Longden, W. McLaren, B. Overduin, B. Pritchard, H. S. Riat, D. Rios, G. Ritchie, M. Ruffier, M. Schuster, D. Sobral, G. Spudich, Y. A. Tang, S. Trevanion, J. Vandrovcova, A. Vilella, S. White, S. P. Wilder, A. Zadissa, J. Zamora, B. L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J. Hubbard, A. Parker, G. Proctor, J. Vogel, and S. M. J. Searle. 2010. Ensembl 2011. *Nucleic Acids Res.* 39:D800–806.
- Gallardo, D., R. N. Pena, M. Amills, L. Varona, O. Ramirez, J. Reixach, I. Díaz, J. Tibau, J. Soler, J. M. Prat-Cuffi, J. L. Noguera, and R. Quintanilla. 2008. Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations in pigs. *Physiol. Genomics* 35:199–209.
- Grindflek, E., J. Szyda, Z. Liu, and S. Lien. 2001. Detection of quantitative trait loci for meat quality in a commercial slaughter pig cross. *Mammalian Genome* 12:299–304.
- Guillou, H., D. Zadavec, P. G. Martin, and A. Jacobsson. 2010. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog. Lipid Res.* 49:186–199.
- Guo, T., J. Ren, K. Yang, J. Ma, Z. Zhang, and L. Huang. 2009. Quantitative trait loci for fatty acid composition in longissimus dorsi and abdominal fat: Results from a White Duroc \times Erhualian intercross F2 population. *Anim. Genet.* 40:185–191.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.
- Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. 462 p. University of Guelph, Guelph, Ontario, Canada.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4:44–57.
- Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37:1–13.
- Ihaka, R., and R. Gentleman. 1996. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299–314.
- Ledur, M. C., N. Navarro, and M. Pérez-Enciso. 2010. Large-scale SNP genotyping in crosses between outbred lines: How useful is it? *Heredity* 105:173–182.
- Mach, N., M. Devant, I. Díaz, M. Font-Furnols, M. A. Oliver, J. A. Garcia, and A. Bach. 2006. Increasing the amount of n-3 fatty acid in meat from young Holstein bulls through nutrition. *J. Anim. Sci.* 84:3039–3048.
- Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, K. Prusa, E. Huff-Lonergan, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig II. Meat and muscle composition. *Mamm. Genome* 12:637–645.
- Meuwissen, T. H., and M. E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* 155:421–430.
- Murakami, M., Y. Taketomi, Y. Miki, H. Sato, T. Hirabayashi, and K. Yamamoto. 2011. Recent progress in phospholipase A2 research: From cells to animals to humans. *Prog. Lipid Res.* 50:152–192.
- Nii, M., T. Hayashi, F. Tani, A. Niki, N. Mori, N. Fujishima-Kanaya, M. Komatsu, K. Aikawa, T. Awata, and S. Mikawa. 2006. Quantitative trait loci mapping for fatty acid composition traits in perirenal and back fat using a Japanese wild boar \times Large White intercross. *Anim. Genet.* 37:342–347.
- Ntawubizi, M., E. Colman, S. Janssens, K. Raes, N. Buys, and S. De Smet. 2010. Genetic parameters for intramuscular fatty acid composition and metabolism in pigs. *J. Anim. Sci.* 88:1286–294.
- Ojeda, A., J. Estellé, J. M. Folch, and M. Pérez-Enciso. 2008. Nucleotide variability and linkage disequilibrium patterns at the porcine FABP5 gene. *Anim. Genet.* 39:468–473.

- Pérez-Enciso, M., A. Clop, J. L. Noguera, C. Ovilo, A. Coll, J. M. Folch, D. Babot, J. Estany, M. A. Oliver, I. Díaz, and A. Sánchez. 2000. A QTL on pig chromosome 4 affects fatty acid metabolism: Evidence from an Iberian \times Landrace intercross. *J. Anim. Sci.* 78:2525–2531.
- Pérez-Enciso, M., and I. Misztal. 2011. Qxpkak.5: Old mixed model solutions for new genomics problems. *BMC Bioinformatics* 12:202.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- Quintanilla, R., R. N. Pena, D. Gallardo, A. Cánovas, O. Ramirez, I. Díaz, J. L. Noguera, and M. Amills. 2011. Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. *J. Anim. Sci.* 89:2963–2971.
- Ramos, A. M., R. P. M. A. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M. S. Hansen, J. Hedegaard, Z.-L. Hu, H. H. Kerstens, A. S. Law, H.-J. Megens, D. Milan, D. J. Nonneman, G. A. Rohrer, M. F. Rothschild, T. P. L. Smith, R. D. Schnabel, C. P. Van Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, and M. A. M. Groenen. 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.
- Rothschild, M. F., and A. Ruvinsky, editors. 2011. *The Genetics of the Pig*, 2nd ed. CABI, Wallingford, Oxfordshire, UK.
- Safran, M., I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel, and D. Lancet. 2010. GeneCards version 3: The human gene integrator. Database (Oxford) 2010.
- Sanchez, M. P., N. Iannuccelli, B. Basso, J. P. Bidanel, Y. Billon, G. Gandemer, H. Gilbert, C. Larzul, C. Legault, J. Riquet, D. Milan, and P. Le Roy. 2007. Identification of QTL with effects on intramuscular fat content and fatty acid composition in a Duroc \times Large White cross. *BMC Genet.* 8:55.
- Sellier, P., L. Maignel, and J. Bidanel. 2010. Genetic parameters for tissue and fatty acid composition of backfat, perirenal fat and longissimus muscle in Large White and Landrace pigs. *Animal* 4:497–504.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100:9440–9445.
- Uemoto, Y., S. Sato, C. Ohnishi, S. Terai, A. Komatsuda, and E. Kobayashi. 2009. The effects of single and epistatic quantitative trait loci for fatty acid composition in a Meishan \times Duroc crossbred population. *J. Anim. Sci.* 87:3470–3476.
- Varona, L., C. Ovilo, A. Clop, J. L. Noguera, M. Pérez-Enciso, A. Coll, J. M. Folch, C. Barragán, M. A. Toro, D. Babot, and A. Sánchez. 2002. QTL mapping for growth and carcass traits in an Iberian \times Landrace pig intercross: Additive, dominant and epistatic effects. *Genet. Res.* 80:145–154.
- Warden, C. H., A. Daluiski, X. Bu, D. A. Purcell-Huynh, C. De Meester, B. H. Shieh, D. L. Puppione, R. M. Gray, G. M. Reaven, Y. D. Chen, and et al. 1993. Evidence for linkage of the apolipoprotein A-II locus to plasma apolipoprotein A-II and free fatty acid levels in mice and humans. *Proc. Natl. Acad. Sci. USA* 90:10886–10890.
- Watkins, P. A., D. Maignel, Z. Jia, and J. Pevsner. 2007. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J. Lipid Res.* 48:2736–2750.
- Wigginton, J. E., and G. R. Abecasis. 2005. PEDSTATS: Descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 21:3445–3447.
- Wong, R. H. F., and H. S. Sul. 2010. Insulin signaling in fatty acid and fat synthesis: A transcriptional perspective. *Curr. Opin. Pharmacol.* 10:684–691.
- Wood, J. D., M. Enser, A. V. Fisher, G. R. Nute, P. R. Sheard, R. I. Richardson, S. I. Hughes, and F. M. Whittington. 2008. Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* 78:343–358.
- Yang, B., N. Navarro, J. L. Noguera, M. Muñoz, T. F. Guo, K. X. Yang, J. W. Ma, J. M. Folch, L. S. Huang, and M. Pérez-Enciso. 2011. Building phenotype networks to improve QTL detection: A comparative analysis of fatty acid and fat traits in pigs. *Journal of Anim. Breed. Genet.* 128:329–343.
- Yang, K. X., J. W. Ma, Y. M. Guo, T. F. Guo, Y. G. Zhao, N. S. Ding, M. Betti, G. S. Plastow, and L. S. Huang. 2010. Correlations between fat depot traits and fatty acid composition in abdominal subcutaneous adipose tissue and longissimus muscle: Results from a White Duroc \times Erhualian intercross F2 population. *J. Anim. Sci.* 88:3538–3545.

PAPER III



Highly accessed

Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition

Yulixaxis Ramayo-Caldas, Nuria Mach, Anna Esteve-Codina, Jordi Corominas,
Anna Castelló, Maria Ballester, Jordi Estellé, Noelia Ibáñez-Escriche, Ana I Fernández,
Miguel Pérez-Enciso and Josep M Folch

BMC Genomics 2012, **13**:547 doi:10.1186/1471-2164-13-547 published online October 11, 2012

RESEARCH ARTICLE

Open Access

Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition

Yulixias Ramayo-Caldas^{1*}, Nuria Mach^{2,3,4}, Anna Esteve-Codina¹, Jordi Corominas¹, Anna Castelló¹, Maria Ballester¹, Jordi Estellé^{2,3,4}, Noelia Ibáñez-Escriche⁵, Ana I Fernández⁶, Miguel Pérez-Enciso^{1,7,8} and Josep M Folch^{1,8}

Abstract

Background: New advances in high-throughput technologies have allowed for the massive analysis of genomic data, providing new opportunities for the characterization of the transcriptome architectures. Recent studies in pigs have employed RNA-Seq to explore the transcriptome of different tissues in a reduced number of animals. The main goal of this study was the identification of differentially-expressed genes in the liver of Iberian x Landrace crossbred pigs showing extreme phenotypes for intramuscular fatty acid composition using RNA-Seq.

Results: The liver transcriptomes of two female groups (H and L) with phenotypically extreme intramuscular fatty acid composition were sequenced using RNA-Seq. A total of 146 and 180 unannotated protein-coding genes were identified in intergenic regions for the L and H groups, respectively. In addition, a range of 5.8 to 7.3% of repetitive elements was found, with SINEs being the most abundant elements. The expression in liver of 186 (L) and 270 (H) lncRNAs was also detected. The higher reproducibility of the RNA-Seq data was validated by RT-qPCR and porcine expression microarrays, therefore showing a strong correlation between RT-qPCR and RNA-Seq data (ranking from 0.79 to 0.96), as well as between microarrays and RNA-Seq ($r=0.72$). A differential expression analysis between H and L animals identified 55 genes differentially-expressed between groups. Pathways analysis revealed that these genes belong to biological functions, canonical pathways and three gene networks related to lipid and fatty acid metabolism. In concordance with the phenotypic classification, the pathways analysis inferred that linolenic and arachidonic acids metabolism was altered between extreme individuals. In addition, a connection was observed among the top three networks, hence suggesting that these genes are interconnected and play an important role in lipid and fatty acid metabolism.

Conclusions: In the present study RNA-Seq was used as a tool to explore the liver transcriptome of pigs with extreme phenotypes for intramuscular fatty acid composition. The differential gene expression analysis showed potential gene networks which affect lipid and fatty acid metabolism. These results may help in the design of selection strategies to improve the sensorial and nutritional quality of pork meat.

Background

Pigs, an important source of human food, accounting for over 40% of the meat produced worldwide. In addition, due to the similarities in anatomy and physiology with humans, they have been used in biomedicine as an important animal model for the study of the genetic basis of metabolic diseases such as obesity, type II

diabetes, metabolic syndrome and atherosclerosis. As well it is often mentioned as the preferred animal species for organ xenotransplantation [1,2].

Over the last decade, a growing awareness of the association between diet and health has led nutritional quality to become a relevant factor in consumers' food choices. A major development has been the recognition that certain fatty acids (FA), such as oleic acid, and α -linolenic acid (ALA), can improve human health status and prevent disease [3,4]. Production of meat with a fatty acids profile more in line with public health recommendations has the

* Correspondence: yulixias.ramayo@uab.es

¹Centre de Recerca en Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra 08193, Spain

Full list of author information is available at the end of the article

potential to improve long-term human health without requiring substantial changes in consumer habits. It is well known that the fatty acid meat composition of pigs is largely dependent on genotype, physiological status and environmental factors such as nutrition [5-11].

The liver a highly specialized organ present in vertebrates and other animals regulates a wide variety of metabolic processes, which play a key role in the digestive function, the decomposition of red blood cells, hormone production and detoxification. Together with adipose tissue and skeletal muscle, the liver is crucial in regulating lipid metabolism. In pigs, the liver is the primary site of *de novo* cholesterol synthesis and fatty acid oxidation, whereas lipogenesis occurs essentially in liver and adipose tissues [12-16].

In the last few years, new high-throughput technologies have been developed for the massive analysis of genomic data. These methodologies yield new opportunities to explore the genetic variability of populations, as well as the characterization of the transcriptome architectures. Until the development of Next-generation sequencing (NGS) technologies, most mRNA expression studies have used microarray or quantitative PCR-based (qPCR) approaches. The development of RNA-Seq, a method based on NGS which consisting of the direct sequencing of RNA molecules present in a given sample, has provided a new tool for both transcriptome characterization and gene expression profiling. In RNA-Seq, the counts corresponding to each transcript can be used for quantification and these sequences can be mapped to the genome for their annotation. In comparison to microarrays, RNA-Seq provides a higher dynamic range, specificity and sensibility [17]. In addition, it provides a picture of the transcriptome, allowing the characterization of alternative splicing, variation in the usage of promoters and polyadenylation sites, non-coding RNAs (ncRNA), single nucleotide variants (SNVs) and transposable elements. Furthermore, RNA-Seq data may allow the discovery of novel transcripts and long intergenic non-coding RNAs (lncRNAs) [18-20].

Recent studies in livestock species have employed RNA-Seq to explore the transcriptome of animal products, such as cow milk [21], bovine embryos [22], and tissue as pig gonads [23], liver, muscle, and abdominal fat [24], sheep bone [25] and bovine abomasal [26]. However, most of the RNA-Seq studies in pigs have included analysis of only a few animals, and ignored within group intrinsic variability. For instance, two single animals of different breeds were compared by Esteve-Codina *et al.*, (2011) and three tissues in two phenotypically extreme full-sib F_2 females formed the basis of Chen *et al.*, (2011) study.

The main goal of this study was the identification of differentially-expressed genes in the liver of groups of Iberian x Landrace crossbred pigs showing extreme

phenotypes for intramuscular fatty acid composition using RNA-Seq. In addition, the porcine hepatic transcriptome was analyzed and transposable elements, new putative protein-coding genes and lncRNAs were identified.

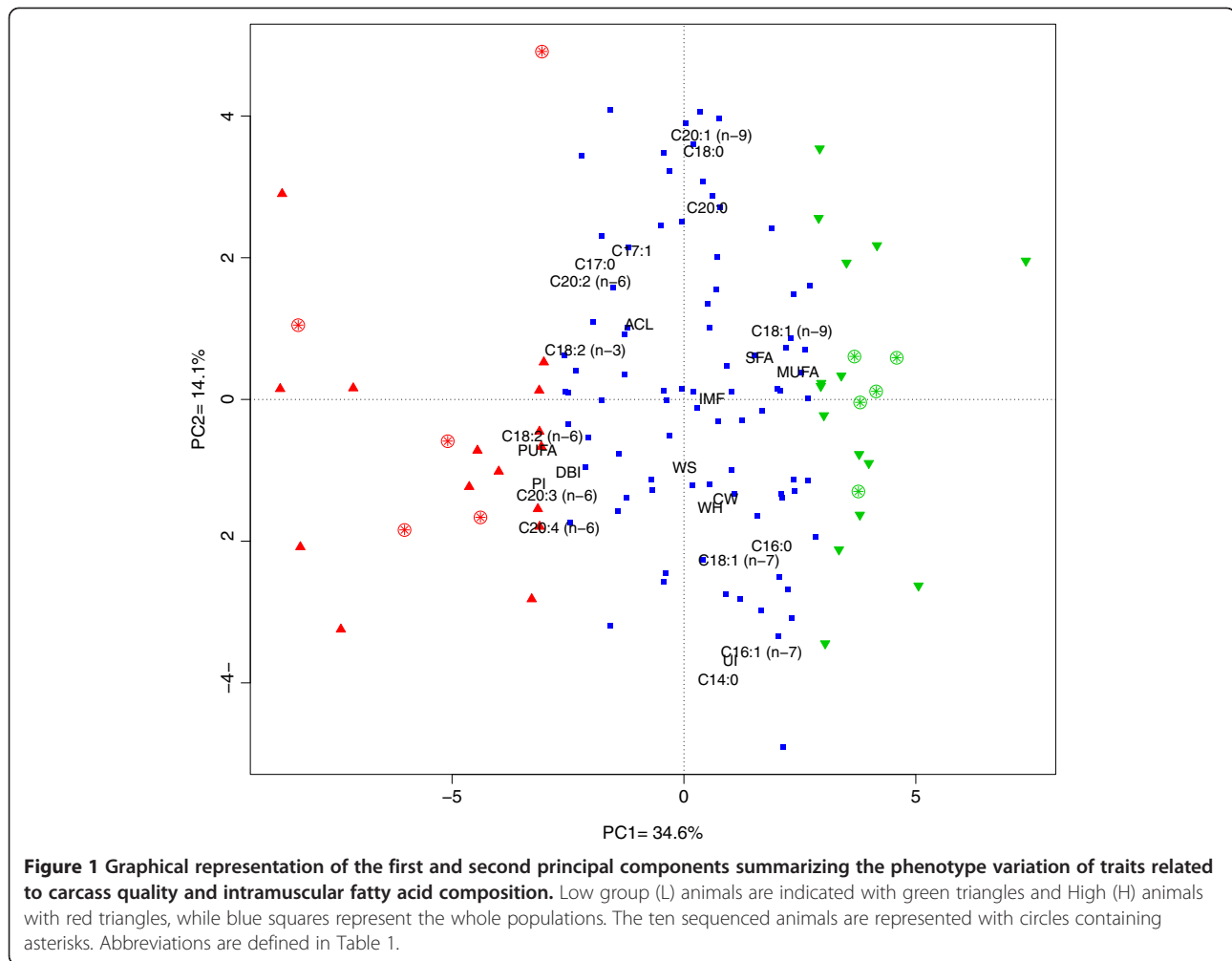
Results and discussion

Phenotypic variation between extreme groups

Analyzed animals were a backcross (25% Iberian x 75% Landrace) obtained by crossing three Iberian (Guadyrbas) boars and 30 Landrace sows. Subsequently five F_1 boards were backcrossed with 26 Landrace sows. A Principal Component Analysis (PCA) was performed to obtain the low-dimensional representation of the data and to describe the phenotypic variation of traits related to carcass quality and intramuscular fatty acid composition. The first two principal components explained the 48.7% of the global phenotypic variance of these traits (PC1=34.6%, PC2=14.1%, Figure 1).

According to the score information for the first principal component the animals were ranked in two groups High (H) and Low (L), of 20 individuals each. Figure 1 shows two clusters of animals with the relative weight of all traits in the two first principal components. The first principal component grouped several traits related to the profile of fatty acids in *Longissimus dorsi* (LD) muscle. Group L showed a higher proportion of saturated (SFA) and monounsaturated fatty acids (MUFA), including palmitoleic and oleic acids. Conversely, H group had a higher content of polyunsaturated acids (PUFA) and related indices like the double bond index (DBI), the unsaturated index (UI) and the peroxidability index (PI). Remarkably, H group also presented a higher proportion of essential PUFA, like linolenic (LA), ALA, eicosadienoic (EDA), eicosatrienoic (ETE) and arachidonic (AA) acids (Table 1). These phenotypic differences are likely determined by genetic variability in: 1) absorption of LA and ALA acids; 2) elongation and desaturation of essential PUFA to longer-chain $\omega-3$ and $\omega-6$ fatty acids; 3) *de novo* synthesis and metabolism of palmitoleic and oleic acids; and 4) transport deposition, storage or degradation and oxidation of all these fatty acids.

Previous studies have reported that, in both backfat and LD muscle, Iberian pigs have higher percentages of palmitic acid, oleic acid, SFA and MUFA, and lower concentrations of LA and ALA acids than commercial breeds [7,27,28]. Moreover, Pascual *et al.*, (2007) [9] reported that Landrace pigs have a higher content of LA and AA acids in their muscle than other commercial breeds. In general, fatter pigs show higher proportions of SFA and MUFA, but less PUFA than lean pigs [6,29]. The genetic architecture of intramuscular FA composition in the Iberian x Landrace backcross was described in a genome-wide association study (GWAS), showing 43 chromosomal intervals associated with these traits



[30]. Since all animals were raised and fed under the same standard management conditions, differences between H and L groups are probably caused by the segregation within the analyzed animals of Iberian and Landrace alleles.

Phenotypic means between groups were compared and significant statistical differences in 73% of the analyzed traits was noted (19/26), mainly relating to intramuscular fatty acid composition (Table 1). The maximum differences between groups were observed for the profiles of essential PUFAs (AA, ETE, LA and ALA acids). Significant differences were also observed for PI and the percentage of palmitic, palmitoleic, heptadecanoic and heptadecanoic acids. From the 20 extreme animals, 10 females were selected for RNA sequencing (five per group). Pedigree information was used to select animals representing the parental genetic diversity. In addition, full-sibs within groups were avoided, animals within groups had different mothers, and four different fathers were selected per group. However, interesting familial relationships between animals of different groups

were retained: there were two pairs of full-sibs and two pair of maternal half-sibs belonging to opposite groups. As before, the phenotypic means differed between groups. However, due to the reduced sample size, only sixteen traits showed significant differences (Additional file 1, Table S1).

Mapping and annotation

The pig liver transcriptome of two groups (H=5, L=5) of phenotypically extreme females for intramuscular fatty acid composition was sequenced. After removal of sequencing adaptors and low-complexity reads, Tophat software was employed [31] to map the reads against the reference pig genome assembly *Scrofa* 9.61. A total of 136.65 M of 100 bp single-end reads (7.28 – 12.43 M of single-end reads per individual) were obtained from two lanes of an *Illumina Hi-Seq 2000* machine. Observed percentages of mapped reads per individual were higher (around 71.42 – 77.75%) than obtained previously in other porcine transcriptome studies; 61.4 - 65.6% [24] and 66.7 [23]. The number of reads and the mean

Table 1 Mean comparison ± standard deviation between high and low groups for the traits included in the principal components analysis (PCA)

Carcass quality	Mean Low	Mean High	Significance
Carcass height (CH)	72.79 ± 9.90	69.57 ± 12.73	NS
Weight of ham (WH)	19.88 ± 2.46	19.66 ± 3.10	NS
Weight of Shoulder (WS)	10.26 ± 1.25	10.93 ± 2.10	NS
Intramuscular fat (IMF)	2.21 ± 0.88	1.49 ± 0.38	NS
Fatty acids			
Saturated FA			
Myristic acid (C14:0)	1.22 ± 0.13	1.12 ± 0.12	NS
Palmitic acid (C16:0)	23.78 ± 0.79	21.39 ± 0.69	***
Heptadecanoic acid (C17:0)	0.20 ± 0.03	0.33 ± 0.09	***
Stearic acid (C18:0)	14.65 ± 1.16	13.69 ± 0.77	*
Arachidic acid (C20:0)	0.26 ± 0.05	0.21 ± 0.08	NS
Monounsaturated FA			
Palmitoleic acid (C16:1 n-7)	2.74 ± 0.24	2.20 ± 0.33	***
Heptadecenoic acid (C17:1)	0.20 ± 0.05	0.32 ± 0.20	**
Oleic acid (C18:1 n-9)	42.57 ± 1.34	34.92 ± 2.96	***
Octadecenoic acid (C18:1 n-9)	4.04 ± 0.27	3.70 ± 0.31	*
Eicosenoic acid (C20:1 n-9)	0.86 ± 0.08	0.77 ± 0.07	NS
Polyunsaturated FA			
Linoleic acid (C18:2 n-6)	7.16 ± 0.52	15.11 ± 1.65	***
α-Linolenic acid (C18:2 n-3)	0.46 ± 0.08	1.10 ± 0.56	***
Eicosadienoic acid (C20:2 n-6)	0.41 ± 0.05	0.63 ± 0.13	***
Eicosatrienoic acid (C20:3 n-6)	0.16 ± 0.03	0.53 ± 0.15	***
Arachidonic acid (C20:4 n-6)	0.84 ± 0.17	3.10 ± 0.77	***
Metabolic ratios			
Average Chain Length (ACL)	17.44 ± 0.02	17.49 ± 0.12	*
Saturated FA (SFA)	40.12 ± 1.51	36.72 ± 1.22	***
Monounsaturated FA (MUFA)	50.72 ± 1.56	42.38 ± 3.13	***
Polyunsaturated FA (PUFA)	9.03 ± 0.64	20.46 ± 2.21	***
Peroxidability index (PI)	12.28 ± 3.80	32.43 ± 4.43	***
Double-bond index (DBI)	0.65 ± 0.20	0.91 ± 0.03	***
Unsaturated index (UI)	1.63 ± 0.50	2.48 ± 0.10	***

NS: p-value > 0.05, * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.

percentages of mapped reads were equivalent for the H and L groups (Table 2).

S-MART [32] was used to calculate the proportion of reads mapping to exons, introns and 1kb upstream/downstream of the annotated genes. As expected, the highest percentage of reads mapped to exons (60.4 – 66.5%), while 11.1 – 16.4% corresponded to introns and the lowest percentage was located either 1 kb upstream or downstream of the annotated genes (4.06 – 5.46%) (Table 3). The proportion of reads mapped to exons of annotated genes was in accordance with the study of Chen *et al.*, (2011) in three pig tissues (60.2 – 74.9%), but was higher than that reported by Esteve-

Table 2 Number of single-end 100 bp reads obtained and percentages of mapped reads per animal

Animal ¹	Total M ²	Mapped M ²	%
L1	9.87	7.28	73.82
L2	12.44	8.88	71.42
L3	13.98	10.62	75.95
L4	11.15	8.47	75.96
L5	14.42	11.21	77.75
H1	16.05	12.43	77.45
H2	12.57	9.77	77.68
H3	14	10.75	76.77
H4	14.94	11.52	77.14
H5	17.23	13.1	76.03
Total	136.65	104.03	76.13

¹ L1 to L5 and H1 to H5 correspond to animals of the L and H groups, respectively.

² Indicate millions of reads.

Codina *et al.*, 2011 (44.1%) in porcine male gonads. These differences (≈ 12%) may be explained by the use of different versions of both the pig genome assembly and annotation, Sscrofa9.61 in the present study and Sscrofa9.58 in Esteve-Codina *et al.*, (2011). Moreover, in the present study a newer version of Tophat was used, which includes improvements in mapping. However, differences between tissues in the proportion of annotated genes cannot be ruled out.

The total number of assembled transcripts with cufflinks was in agreement with the previously reported pig liver transcriptome [24]. These transcripts fall into the following categories: annotated exons (8.7 – 11%), intron retention events (11 – 13.5%), intergenic transcripts (19.1 – 21.7%), potentially novel isoforms of genes (17.1 – 20.3%), known isoforms (14.7 – 17.8%), pre-mRNA molecules (2 – 3.3%) and polymerase run-on fragments (5.9 – 8.3%) (Table 4).

Table 3 Proportion of reads mapping to exons, introns or within 1 Kb upstream or downstream of the annotated genes

Animal ¹	% Exons	% Introns	% _5' or _3' ²
L1	60.44	16.44	5.03
L2	66.48	13.59	4.06
L3	65.52	13.06	5.31
L4	63.35	14.71	4.57
L5	62.86	12.71	4.6
H1	63.54	13.52	5.44
H2	62.1	15.08	4.78
H3	64.87	14.11	4.28
H4	65.95	11.12	5.46
H5	66.46	12.98	4.44

¹ L1 to L5 and H1 to H5 correspond to animals of the L and H groups, respectively.

² Reads located either 1 Kb upstream or downstream of the annotated genes.

Table 4 Number of transcripts assembled (TA) with Cufflinks and the percentage they represent in each sample

Code	L1		L2		L3		L4		L5		H1		H2		H3		H4		H5	
	TA	%	TA	%	TA	%	TA	%	TA	%	TA	%	TA	%	TA	%	TA	%	TA	%
=	1971	8.7	2486	10.4	2526	11	2244	9.2	2762	10.2	2642	10.1	2363	9.4	2761	9.9	2119	10.4	2838	10.5
c	4059	17.8	4202	17.6	3953	17.2	4203	17.3	4289	15.9	3955	15.1	4083	16.3	4183	15.0	3569	17.5	3961	14.7
e	745	3.3	508	2.1	518	2.3	672	2.8	576	2.1	616	2.4	736	2.9	604	2.2	489	2.4	529	2.0
i	3115	13.7	2719	11.4	2629	11.4	3153	13.0	3248	12.0	3301	12.6	3373	13.5	3637	13.1	2250	11.0	3234	12.0
j	4348	19.1	4709	19.7	4436	19.3	4666	19.2	5373	19.9	5288	20.2	4833	19.3	5658	20.3	3488	17.1	5428	20.1
o	1067	4.7	1175	4.9	1139	5.0	1157	4.8	1237	4.6	1142	4.4	1154	4.6	1170	4.2	1129	5.5	1204	4.5
p	1570	6.9	1656	6.9	1566	6.8	1682	6.9	1635	6.1	1603	6.1	1606	6.4	1676	6.0	1699	8.3	1598	5.9
s	60	0.3	77	0.3	72	0.3	76	0.3	94	0.4	86	0.3	71	0.3	72	0.3	68	0.3	100	0.4
u	4447	19.5	4560	19.1	4541	19.7	4789	19.7	5468	20.2	5456	20.9	5089	20.3	5680	20.4	4421	21.7	5638	20.9
x	1392	6.1	1838	7.7	1630	7.1	1663	6.8	2346	8.7	2066	7.9	1761	7.0	2415	8.7	1192	5.8	2496	9.2
Total	22774	100	23930	100	23010	100	24305	100	27028	100	26155	100	25069	100	27856	100	20424	100	27026	100

Class codes described by Cuffcompare: "=" Exactly equal to the reference annotation, "c" Contained in the reference annotation, "e" possible pre-mRNA molecule, "i" An exon falling into an intron of the reference, "j" New isoforms, "o" Unknown, generic overlap with reference, "p" Possible polymerase run-on fragment, "s" An intron of the transfrag overlapping a reference intron on the opposite strand, "u" Unknown, intergenic transcript, "x" Exonic overlap with reference on the opposite strand. L1 to L5 and H1 to H5 correspond to animals of the L and H groups, respectively.

Gene expression analysis

The total amount of expressed genes in liver was similar between groups (L= 8797 – 10161, H= 8765 – 10083). Taking into account only those genes with a mean FPKM (normalized number of fragments per kilobase of exon per million reads) higher than zero, an aggregate of 10,485 expressed genes in L and 10,626 in H groups was observed. A total of 10,280 common genes were expressed in both groups. The correlation of mean gene expression levels between both groups (H vs L) was very high ($r = 0.99$), suggesting that the major fraction of the liver transcriptome is conserved between groups. Gene expression distribution reveals that less than 10% of these genes were expressed between 1 – 10 FPKM; around 42% between 10 FPKM - 100 FPKM; 38% among 100 – 1000 FPKM and, approximately, 8% more than 1000 FPKM (Additional file 2, Figure S1).

All 10 individuals were also assayed with the Gene-Chip® Porcine microarray (*Affymetrix*, Santa Clara, CA) which allows the expression analysis of 20,201 *Sus scrofa* genes. After probe normalization, correlation between the expression data of microarrays and RNA-Seq was calculated. In accordance with previous studies [17,23], a strong Spearman correlation ($r=0.72$) was observed (Additional file 3, Figure S2). Results from both technologies were, in general, more similar for genes that showed intermediate expression values, whereas major differences were observed for low and high expressed genes in the *Affymetrix* microarray data. This pattern can be explained by the higher dynamic range of RNA-Seq [17,33]. Finally, in line with the previous description of liver transcriptome [34], the top 100 expressed genes showed an overrepresentation in biological gene ontologies related to oxidoreductase activity, transport,

proteolysis, translation, signal transduction, cholesterol homeostasis and lipid transport ($p < 0.001$).

Transposable elements analysis

The percentage of repetitive elements identified in the pig liver transcriptome was around 5.8 – 7.3% (Additional file 4, Table S2), similar to that found in male gonad transcriptome (7.3%) [23]. However, it should be noted that the total length of base pairs masked and the total number of transcripts were higher in male gonad transcriptome [23] than in our study. Two possible explanations may account for these differences: 1) in liver from 7.3 – 12.4 M of single-end reads per individual were obtained (Table 2), whereas in gonads [23] a total of 20 M paired-end reads were observed and, thus, a better fragment distribution and a higher number of transcripts were analysed; 2) the transcriptome complexity has been reported [35] to be higher in kidney, testes and brain tissues in comparison to liver and muscle.

Gene orthology and lncRNAs detection

From the unannotated (Sscrofa 9.61 assembly) intergenic expressed regions, a range of 3488–5658 intergenic novel transcripts were identified in each sample (Table 4). However, to find not annotated genes, transcripts expressed in at least four of the five animals of each group were considered. Then, Augustus software [36] was used to examine which of these transcripts were predicted to encode proteins. A total of 146 and 180 putative proteins were identified for the L and H groups, respectively (Table 5). According to BLASTP results, these proteins correspond for L group to: 19 novel computationally predicted and 95 known human proteins, 3 novel and 107 known bovine proteins and 43 novel and 7 known porcine proteins.

Table 5 Putative proteins identified in each H and L groups and orthologies detected against *Homo sapiens*, *Bos taurus* and *Sus scrofa* protein databases

Group	Total	<i>Bos taurus</i>	<i>Homo sapiens</i>	<i>Sus scrofa</i>	Esteve-Codina et al. (2011)
L	146	110	114	50	82
H	180	133	135	59	104

Predicted proteins reported in pigs by Esteve-Codina et al., (2011) are also included.

Similarly, for the H group predicted proteins correspond to: 25 novel computationally predicted and 110 known human proteins, 5 novel and 128 known bovine proteins and 51 novel and 8 known porcine proteins. Interestingly, in both H and L groups, around 86% of the predicted novel proteins were found in the *Sus scrofa* protein database [ftp://ftp.ensembl.org/pub/release-65/fasta/sus_scrofa/pep/] as novel computationally predicted proteins. Moreover, the number of matches increased to 58% (104/180) when the predicted novel proteins were compared against the putative coding transcripts reported by Esteve-Codina *et al.*, (2011). This result indicates that a high number of genes are not annotated in the *Sscrofa* 9.61 pig genome assembly and they are expressed in both liver and gonad tissues. Finally, these results constitute an experimental confirmation of the novel computationally predicted genes in pigs.

For lncRNAs annotation, the previously reported sequences in pig male gonad transcriptome [23] were used as a reference database. A total of 186 (L) and 270 (H) of these putative lncRNAs was also expressed in pig liver. Within groups, 101 and 108 lncRNAs were expressed in all L and H animals, respectively, but only 89 lncRNAs were expressed in both groups (Additional file 5, Figure S3).

Differential gene expression analysis

DESeq software [37] was employed to detect differentially-expressed (DE) genes between H and L groups. First, some exploratory analyses to estimate the variance and quality of the data were performed. Per-gene estimates of the base variance against the base levels showed that the fit (red line) closely followed the single-gene estimates (Additional file 6, Figure S4). The 'residualsEcdfPlot' function, which checks the uniformity of the cumulative probabilities, revealed a similar pattern for the curves of the empirical cumulative density functions (ECDF) in both groups (Additional file 7, Figure S5). It was also noted that ECDF followed well the diagonal, except for very low counts, but that is to be expected because at this level shot noise dominates and, therefore, the deviations become stronger (Additional file 8, Figure S6). Afterwards DE analysis between groups was performed. Figure 2 shows that, at selected cut-off ($-\log_{10}(p\text{-value}) > 2.3$ or $q\text{-value} \leq 0.17$), there is a clear departure from

expected among transcripts accepted as differentially expressed (indicated by the blue trend being above the straight red line). Therefore according to the employed cut-off (fold change ≥ 1.5 and $p\text{-value} \leq 0.005$), 55 protein-coding genes, two pseudogenes [ENSEMBL_Id: ENSSSCG00000004170, ENSSSCG00000016238] and one non-coding RNA [ENSEMBL_Id: ENSSSCG00000019010] were identified as differentially-expressed (Figure 3, Table 6).

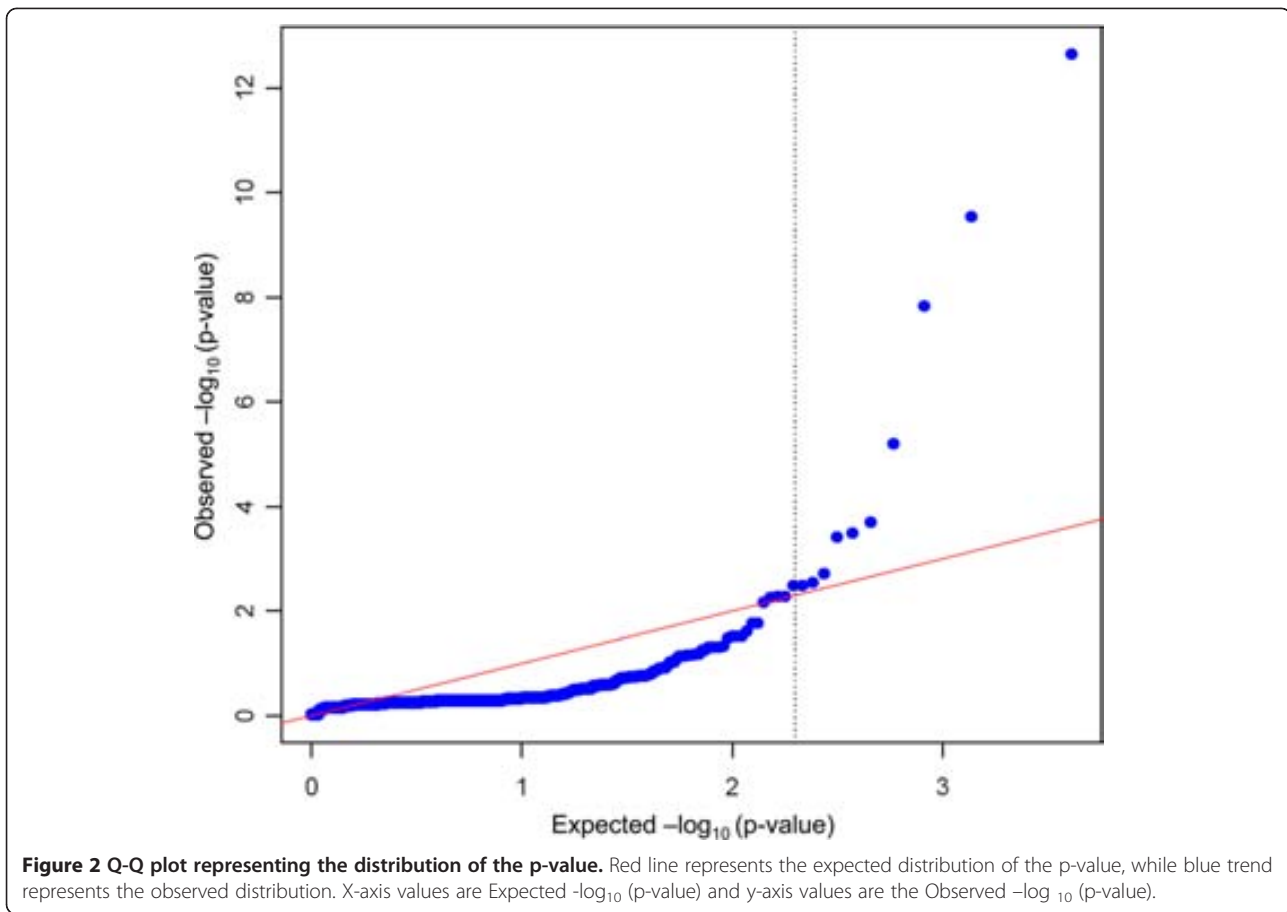
In order to validate the expression data obtained by RNA-Seq, five genes (*APOA2*, *LPIN1*, *ME3*, *CYP7A1* and *CYP2C49*) were selected among the differentially-expressed protein-coding genes to perform real time reverse transcription (RT-qPCR) assays. When the pattern of gene expression levels was compared, strong correlations ranking from 0.79 to 0.96 between RT-qPCR and RNA-Seq platforms were observed, confirming the high reproducibility of the data (Additional file 9, Table S3).

Interestingly, one of the studied genes, the *CYP2C49* [ENSEMBL_Id: ENSSSCG00000010488] which belongs to the highly diverse superfamily of CYP450 [38] and it is homologue to the human *CYP2C9* gene, was located in a genomic region in which copy number variation (CNV) has been previously described in pigs [39]. In order to assess whether observed differences of gene expression were influenced by differences in the CNV between animals, a real time quantitative PCR (qPCR) to determine the number of copies of the *CYP2C49* gene was developed. For the first time, CNV affecting the *CYP2C49* gene was described with relative quantification values ranging from 1 to 5.2 copies (Additional file 9, Table S3). However, no correlation between the number of copies and gene expression was observed. Therefore, further analysis will be necessary to elucidate the possible role of these structural variants in the fatty acid metabolism.

Moreover, it is noteworthy that six of the differentially-expressed genes related to fatty acids metabolism in liver (*APOB*, *CYP7A1*, *APOA2*, *THBS1*, *THEM5*, *ME3*) were previously reported as associated with the profile of intramuscular fatty acid composition in a GWAS study in the same animal population [30]. Therefore, they can be considered as interesting candidate genes and this suggest their role in the fatty acid metabolism processes in pigs in both liver and IMF tissues (Table 7).

Functional clustering of differentially-expressed genes in the liver

From the 55 differentially-expressed protein-coding genes, 26 were up-regulated and 29 were down-regulated in H group in comparison to L (Table 6). To gain insight into the liver tissue processes that differed between groups, the list of the differentially-expressed genes was explored using the core analysis function



included in Ingenuity Pathways Analysis (IPA). Initially, the pig gene IDs were converted to human genes but five protein-coding genes did not match with human homologs [Ensembl Ids: ENSSSCG00000007873, ENSSSCG00000003971, ENSSSCG00000014368, ENSSSCG00000013116, ENSSSCG00000001229], and therefore only 50 pig genes were eligible for network construction.

The top seven biological functions identified by IPA included categories related to a wide variety of physiological and biological events, such as lipid metabolism, small molecule biochemistry, molecular transport, drug metabolism, energy production, nucleic acid metabolism, and vitamin and mineral metabolism (Table 8). A specific examination of the lipid metabolism IPA molecular and cellular function revealed that most of the transcripts relating to lipid metabolism were up-regulated in H group compared to L group. Remarkably, genes that play a crucial role in lipoprotein synthesis (*APOB*), cholesterol metabolism (*ABCG8*, *CYP2C9*, *CYP2C19*, *CYP4A11*, *APO* and *CYP7A1*), oxidation of lipids and palmitic fatty acids (*MTMR7*), and induction of lipogenic gene transcription (*LPINI*) were up-regulated in H group in contrast to L group. On the contrary, genes involved with accumulation of triacylglycerol (*AQP7*),

uptake of lipids and myristic acid (*THBS1*), and fatty acids biosynthesis (*ME3*) were down-regulated in H group as opposed to L group. The malic enzyme is involved in supplying NADPH for the reductive biosynthesis of fatty acids [40]. Based on the above observations it is tempting to speculate that ALA and LA acids reaching liver tissue inhibit the expression of *ME3* gene, and consequently, at least partially, reduce lipogenic activity. This is in agreement with Guillevic et al. [8], who reported that ALA acid enriched diets decreased malic enzyme activity in liver and subcutaneous adipose tissue of pigs.

Interestingly, the most representative canonical pathways significantly modulated in liver when comparing H vs L groups were involved in Endotoxin lipopolysaccharide / pro-inflammatory cytokines (LPS/IL-1) mediated inhibition of retinoid X receptors (RXR) function, fatty acid metabolism (including AA and LA acids) and pregnane X receptor / farnesoid X receptor (FXR/RXR) activation (Figure 4 and Table 9), in which the up-regulation of *ABCG8*, *APOB* and *CYP7A1* genes was observed. Likewise, the present findings underscore that H group increased the expression of gene sets regulated by peroxisome proliferator-activated receptors alpha (PPAR- α)

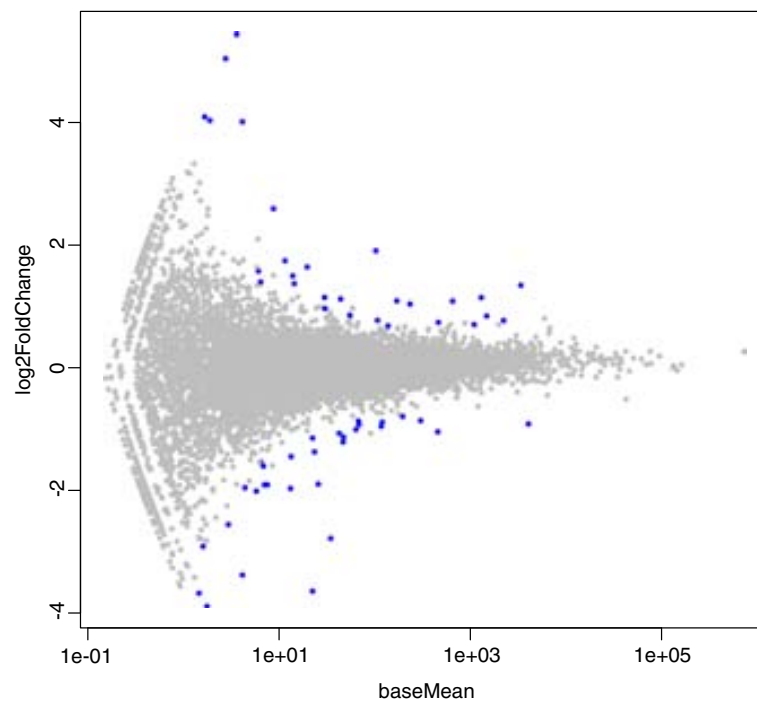


Figure 3 Plot of the 55 differentially-expressed protein-coding genes (represented in blue) with fold change ≥ 1.5 and p -value ≤ 0.005 . X-axis values are base mean expression values and y-axis values are the \log_2 (fold change).

(*APOA2*, *CYP2C9*, *CYP2C19*) and RXR (*ABCG8*, *CYP7A1*) transcription factors, both of which are shown to have an important role in lipid homeostasis. For instance, PPAR- α is an important regulator of cellular fatty acid uptake and intracellular fatty acid transport, mitochondrial and peroxisomal fatty acid oxidation, ketogenesis, and gluconeogenesis in several species [41-43], whereas RXR plays a crucial role in the transcriptional regulation of a spectrum of genes controlling cholesterol homeostasis and bile acid homeostasis, together with nuclear receptor FXR, a key transcription factor that regulates cholesterol 7 α -hydroxylase (*CYP7A1*) activity and mRNA levels.

In addition, the up-regulation of PPAR- α and RXR were coupled with the increased expression of lipin (*LPIN1*) and *CYP7A1* genes. In mice, it has been reported that *LPIN1* selectively activates a subset of coactivator 1 α (*PGC-1 α*) target pathways involved in fatty acid oxidation and mitochondrial oxidative phosphorylation, while suppressing the lipogenic program and lowering circulating lipid levels [44]. Lipin activates mitochondrial fatty acid oxidative metabolism by inducing expression of the nuclear receptor PPAR- α , a known *PGC-1 α* target, and via direct physical interactions with PPAR- α and *PGC-1 α* . Furthermore, *CYP7A1* has been shown to be a key factor of hepatic cholesterol homeostasis. All together these results suggest that H group may present greater uptake of fatty acids into hepatocytes (mainly LA and ALA acids). It is likely that the

higher PUFA bioavailability in liver may affect expression of PPAR- α , RXR and their target genes, inducing a greater stimulation of both peroxisomal and mitochondrial β -oxidation, and leading to reduced triglyceride and cholesterol synthesis, and an enhanced elimination of cholesterol from the liver via bile acid formation. This intriguing possibility remains to be demonstrated, although there is evidence that FA, in particular unsaturated FA, exert many of their biological effects by regulating the activity of numerous transcription factors in liver, such as PPAR- α [45]. Recently, [46] has demonstrated that FA oxidation is regulated by hepatic MUFA to PUFA ratio through the activation of PPAR- α . In agreement to our results, hepatic expression of PPAR- α was higher in pigs fed with a higher level of PUFA. This is also in line with the lower IMF content in H group than in L animals, and the lower proportion of *de novo* fatty acids in the IMF. Therefore, these transcriptome changes may reflect counter mechanisms of liver tissue to respond or compensate for changes in IMF fatty acid profile, which depends on possible metabolism of FAs and the possibility of being synthesized by the pig adipose tissue [47]. However, the question remains how different types of FA control the expression of genes and a direct examination of the effect of each individual FA on porcine muscle fatty acid composition is needed.

Finally, the identified genes were mapped to three genetic networks. The first, having an IPA network score of 38 and 16 focus genes, presented functions related to Lipid

Table 6 Description of the differentially-expressed genes detected between High and Low groups with fold change ≥ 1.5 and p-value ≤ 0.005

Ensembl_Gene_Id	Human homolog	Fold change	p-value	q-value	Gene biotype
ENSSSCG00000010610	GSTO1	3.8	1.1×10^{-16}	2.2×10^{-13}	protein_coding
ENSSSCG00000010992	AQP7	-6.7	3.0×10^{-13}	2.9×10^{-10}	protein_coding
ENSSSCG00000016401	KIF1A	-12.5	2.2×10^{-11}	1.4×10^{-8}	protein_coding
ENSSSCG00000010488	CYP2C9	2.5	1.3×10^{-8}	6.3×10^{-6}	protein_coding
ENSSSCG00000013865	NWD1	16.2	5.2×10^{-7}	2.0×10^{-4}	protein_coding
ENSSSCG00000012015	C21orf91	20.4	1.0×10^{-6}	3.0×10^{-4}	protein_coding
ENSSSCG00000007873	-	-3.7	1.4×10^{-6}	4.0×10^{-4}	protein_coding
ENSSSCG00000009871	SDS	2.2	7.8×10^{-6}	2.0×10^{-3}	protein_coding
ENSSSCG00000002383	FOS	-2.0	1.3×10^{-5}	3.0×10^{-3}	protein_coding
ENSSSCG00000019010	-	-14.3	1.6×10^{-5}	3.0×10^{-3}	snRNA
ENSSSCG00000000044	C22orf32	-3.8	1.8×10^{-5}	3.0×10^{-3}	protein_coding
ENSSSCG00000003891	CYP4A11	2.1	3.2×10^{-5}	1.0×10^{-2}	protein_coding
ENSSSCG00000016238	-	3.1	3.5×10^{-5}	1.0×10^{-2}	pseudogene
ENSSSCG00000011937	MORC1	6.1	3.9×10^{-5}	1.0×10^{-2}	protein_coding
ENSSSCG00000010487	CYP2C19	2.1	5.1×10^{-5}	1.0×10^{-2}	protein_coding
ENSSSCG00000006614	THEM5	18.0	1.4×10^{-4}	2.0×10^{-2}	protein_coding
ENSSSCG00000005385	NR4A3	-2.7	1.4×10^{-4}	2.0×10^{-2}	protein_coding
ENSSSCG00000006580	S100A2	-3.7	2.2×10^{-4}	2.0×10^{-2}	protein_coding
ENSSSCG00000000231	ANKRD33	-2.6	3.0×10^{-4}	3.0×10^{-2}	protein_coding
ENSSSCG00000001642	TBCC	-14.3	3.2×10^{-4}	3.0×10^{-2}	protein_coding
ENSSSCG00000003971	-	-2.0	3.3×10^{-4}	3.0×10^{-2}	protein_coding
ENSSSCG00000015294	CR1	-2.0	3.7×10^{-4}	3.0×10^{-2}	protein_coding
ENSSSCG00000006238	CYP7A1	2.1	5.6×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000004789	THBS1	-1.7	6.0×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000012832	MXRA5	-2.1	6.5×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000004946	ZWILCH	2.2	6.7×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000007888	TNFRSF17	-3.7	6.7×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000008595	APOB	1.6	7.9×10^{-4}	5.0×10^{-2}	protein_coding
ENSSSCG00000014919	ME3	-2.3	8.4×10^{-4}	6.0×10^{-2}	protein_coding
ENSSSCG00000014368	-	-2.0	1.0×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000007529	SYCP2L	2.6	1.0×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000013116	-	16.4	1.1×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000004052	FNDC1	-1.8	1.2×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000002277	SPTB	-4.0	1.2×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000016645	C7orf53	-7.7	1.3×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000008203	IGKV2-40	-1.8	1.3×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000001229	-	-2.0	1.3×10^{-3}	7.0×10^{-2}	protein_coding
ENSSSCG00000000151	APOL6	3.4	1.5×10^{-3}	8.0×10^{-2}	protein_coding
ENSSSCG000000004170	-	2.2	1.9×10^{-3}	8.0×10^{-2}	pseudogene
ENSSSCG00000016190	SLC11A1	-3.8	1.9×10^{-3}	8.0×10^{-2}	protein_coding
ENSSSCG00000006355	APOA2	1.8	1.9×10^{-3}	8.0×10^{-2}	protein_coding
ENSSSCG00000015747	MYOM2	2.8	2.3×10^{-3}	9.0×10^{-2}	protein_coding
ENSSSCG00000014824	RELT	-12.5	2.7×10^{-3}	9.0×10^{-2}	protein_coding

Table 6 Description of the differentially-expressed genes detected between High and Low groups with fold change ≥ 1.5 and p-value ≤ 0.005 (Continued)

ENSSSCG000000008455	ABCG8	17.1	2.7×10^{-3}	9.0×10^{-2}	protein_coding
ENSSSCG000000003777	SLC44A5	2.6	2.8×10^{-3}	9.0×10^{-2}	protein_coding
ENSSSCG000000003999	A1BG	-5.9	2.8×10^{-3}	9.0×10^{-2}	protein_coding
ENSSSCG000000002375	RPS6KL1	-2.2	3.2×10^{-3}	1.3×10^{-1}	protein_coding
ENSSSCG000000001006	TUBB2B	-2.9	3.4×10^{-3}	1.4×10^{-1}	protein_coding
ENSSSCG000000007478	ATP9A	-1.8	3.5×10^{-3}	1.4×10^{-1}	protein_coding
ENSSSCG000000009992	UQCR10	1.8	3.9×10^{-3}	1.6×10^{-1}	protein_coding
ENSSSCG000000010829	MOSC1	1.7	4.0×10^{-3}	1.6×10^{-1}	protein_coding
ENSSSCG000000017923	ALOX15	-3.0	4.3×10^{-3}	1.6×10^{-1}	protein_coding
ENSSSCG000000002847	GPT2	1.7	4.6×10^{-3}	1.7×10^{-1}	protein_coding
ENSSSCG000000004787	GPR176	3.0	4.7×10^{-3}	1.7×10^{-1}	protein_coding
ENSSSCG000000006985	MTMR7	1.7	4.8×10^{-3}	1.7×10^{-1}	protein_coding
ENSSSCG000000000709	PLEKHG6	-1.9	5.0×10^{-3}	1.7×10^{-1}	protein_coding
ENSSSCG000000010892	KCNT2	2.0	5.0×10^{-3}	1.7×10^{-1}	protein_coding
ENSSSCG000000008624	LPIN1	1.6	5.0×10^{-3}	1.7×10^{-1}	protein_coding

Metabolism, Small Molecule Biochemistry and Vitamin and Mineral Metabolism (Figure 5). The second, with a score of 23 and 11 focus genes centred on Lipid Metabolism, Molecular Transport and Small Molecule Biochemistry (Additional file 10, Figure S7), and the third network scoring 21 and 10 focus genes was associated with Carbohydrate Metabolism, Lipid Metabolism and Molecular Transport (Additional file 11, Figure S8). When the top three IPA networks were merged a connection between them was observed (Figure 6), suggesting that the differentially-expressed genes detected in this study are linked and play an important role in lipid metabolism. Remarkably, IPA results are in conformity with the design of this experiment, which inferred that LA and AA acids metabolism were altered between the groups of sequenced individuals.

Implications

In the present study, RNA-Seq was used for the analysis of the pig liver transcriptome in animals of extreme phenotypes for intramuscular fatty acid composition.

The liver plays an important role in lipid metabolism and, thus, the analysis of liver transcriptome in extreme pigs for intramuscular fatty acid composition may be relevant to elucidate its functional complexity. Although the main goal of this study was to find differentially-expressed genes between phenotypically extreme animals, the use of RNA-Seq allowed the identification of transposable elements, lncRNAs and new protein-coding genes in the porcine liver transcriptome.

The first principal component of PCA analysis classified animals in two extreme groups for the fatty acid composition of LD muscle. Group H of animals had a higher content of PUFA, including essential FA such as LA, ALA, ELE and AA acids than group L animals. Conversely, the latter had a higher content of SFA and MUFA, palmitoleic and oleic acids.

The lipid content and fatty acid profile of muscle plays an important role in the tenderness, flavour and juiciness of cooked meat [6]. In swine production, the reduction of intramuscular fat (IMF) in some breeds due to a preference selection for lean pigs, has affected

Table 7 Differentially-expressed genes previously reported to be associated with the profile of intramuscular fatty acid composition in a genome-wide association study

Ensembl gene ID	Chr	Start (bp) ¹	End (bp) ¹	Gene name	Fatty acid
ENSSSCG000000004789	1	138129409	138145238	THBS1	C18:1(n-9), C18:2(n-6), MUFA
ENSSSCG000000008595	3	109052838	109076900	APOB	C16:1(n-7), ratio C16:1(n-7)/c16:0
ENSSSCG000000006238	4	77173363	77202771	CYP7A1	C16:1(n-7), C18:2(n-6)
ENSSSCG000000006355	4	92745976	92747590	APOA2	C16:0, C18:2(n-6), rate MUFA/SFA
ENSSSCG000000006614	4	101212520	101222091	THEM5	rate MUFA/SFA
ENSSSCG000000014919	9	20546397	20647476	ME3	rate C20:1/C20:0

¹ The genomic coordinates are expressed in bp and are relative to the *Sus scrofa* April 2009 genome sequence assembly (Sscrofa9).

Table 8 Description of the top seven molecular and cellular biological functions significantly modulated in the liver tissue when comparing H relative to L animals

Category	Genes	p-value
Lipid Metabolism	<i>ABCG8, ALOX15, AQP7, APOB, CYP2C9, THBS1, APOA2, ME3, NR4A3, LPIN1, CYP7A1, MTMR7, CYP2C19, CYP4A11</i>	1.15×10^{-7}
Small Molecule Biochemistry	<i>ABCG8, ALOX15, AQP7, APOB, CYP2C9, THBS1, APOA2, ME3, GSTO1, FOS, NR4A3, LPIN1, SPTB, CYP7A1, MTMR7, CYP2C19, SDS, SLC11A1, CYP4A11</i>	1.15×10^{-7}
Molecular Transport	<i>ABCG8, AQP7, ALOX15, APOB, THBS1, APOA2, GSTO1, FOS, NR4A3, LPIN1, CYP7A1, SLC11A1, FNDC1</i>	2.62×10^{-6}
Drug Metabolism	<i>FOS, CYP2C9, THBS1, CYP2C19</i>	7.20×10^{-6}
Energy Production	<i>NR4A3, LPIN1, CYP2C9, APOA2, ME3, CYP2C19, SDS, CYP4A11</i>	7.20×10^{-6}
Nucleic Acid Metabolism	<i>CYP2C9, THBS1, CYP2C19</i>	5.85×10^{-6}
Vitamin and Mineral Metabolism	<i>ABCG8, APOB, CYP2C9, APOA2, CYP7A1, CYP2C19, GSTO1</i>	8.65×10^{-6}

Statistical significance of pathway modulation was calculated via a right-tailed Fisher's Exact test in Ingenuity Pathway and represented as $-\log(P \text{ value})$: $-\log$ values exceeding 1.30 were significant false discovery rate (FDR) < 0.05.

meat quality. From this point of view, PUFA has a negative effect on the oxidative stability of muscle, which, in turn, affects flavour and muscle colour [6].

Table 9 Description of the top six canonical pathways significantly modulated in liver tissue when comparing H to L animals

Ingenuity canonical pathways	Genes
LPS/IL-1 Mediated Inhibition of RXR Function	<i>ABCG8, CYP2C9, CYP7A1, CYP2C19, CYP4A11, GSTO1</i>
Arachidonic Acid Metabolism	<i>ALOX15, CYP2C9, CYP2C19, CYP4A11</i>
Fatty Acid Metabolism	<i>CYP2C9, CYP2C19, SDS, CYP4A11</i>
PXR/RXR Activation	<i>CYP2C9, CYP7A1, CYP2C19</i>
Linoleic Acid Metabolism	<i>ALOX15, CYP2C9, CYP2C19</i>
FXR/RXR Activation	<i>ABCG8, APOB, CYP7A1</i>

Statistical significance of pathway modulation was calculated via a right-tailed Fisher's Exact test in Ingenuity Pathway analysis and represented as $-\log(P \text{ value})$: $-\log$ values exceeding 1.30 were significant FDR < 0.05.

On the other hand, desirable sensorial characteristics tend to be associated with MUFA and SFA [6,48,49]. Lipid and fatty acid compositions of food have an important impact on human health, with a high consumption of SFA associated with obesity, high plasma cholesterol and cardiovascular disease [50,51]. Conversely, PUFAs, mainly $\omega-3$, have been considered beneficial for human health, by reducing serum low-density lipoprotein-C, total cholesterol concentration and modulating immune functions and inflammatory processes [52-54].

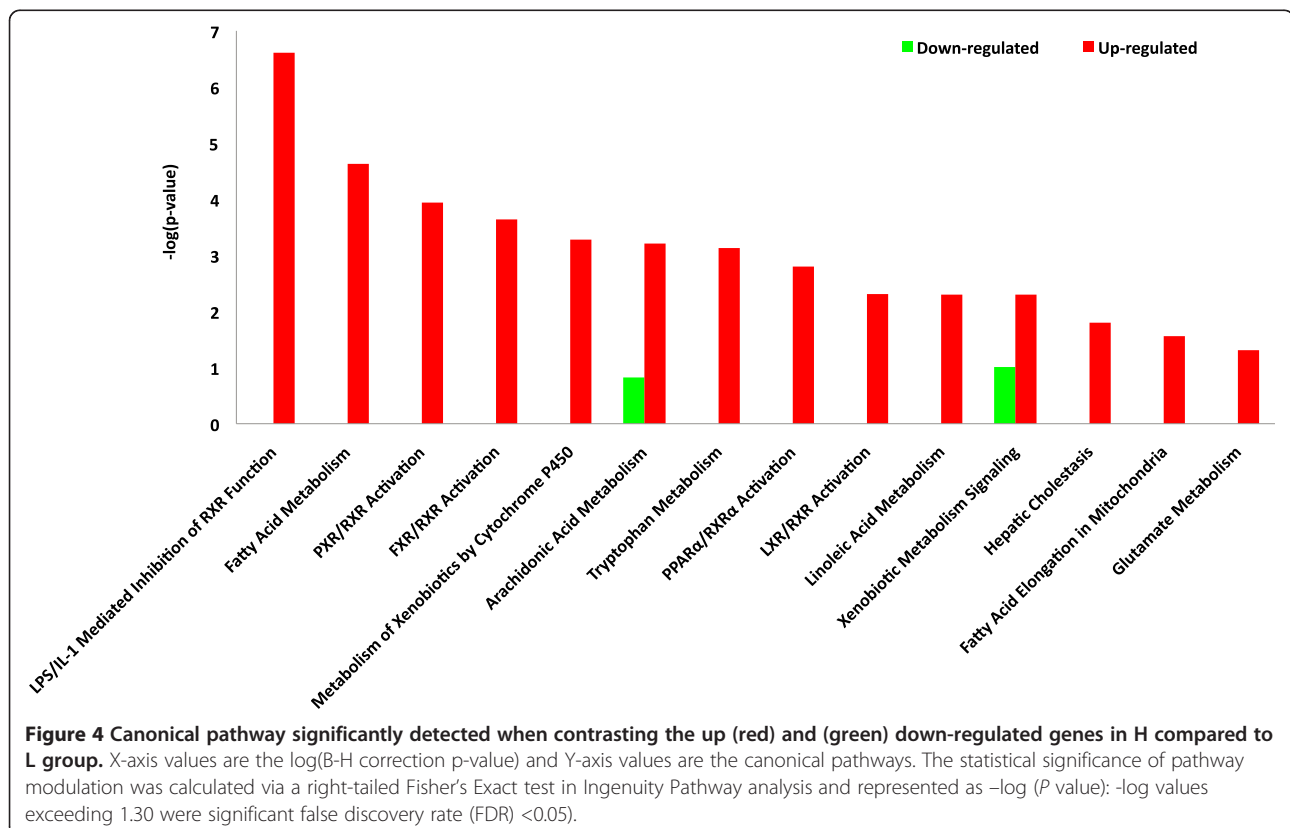
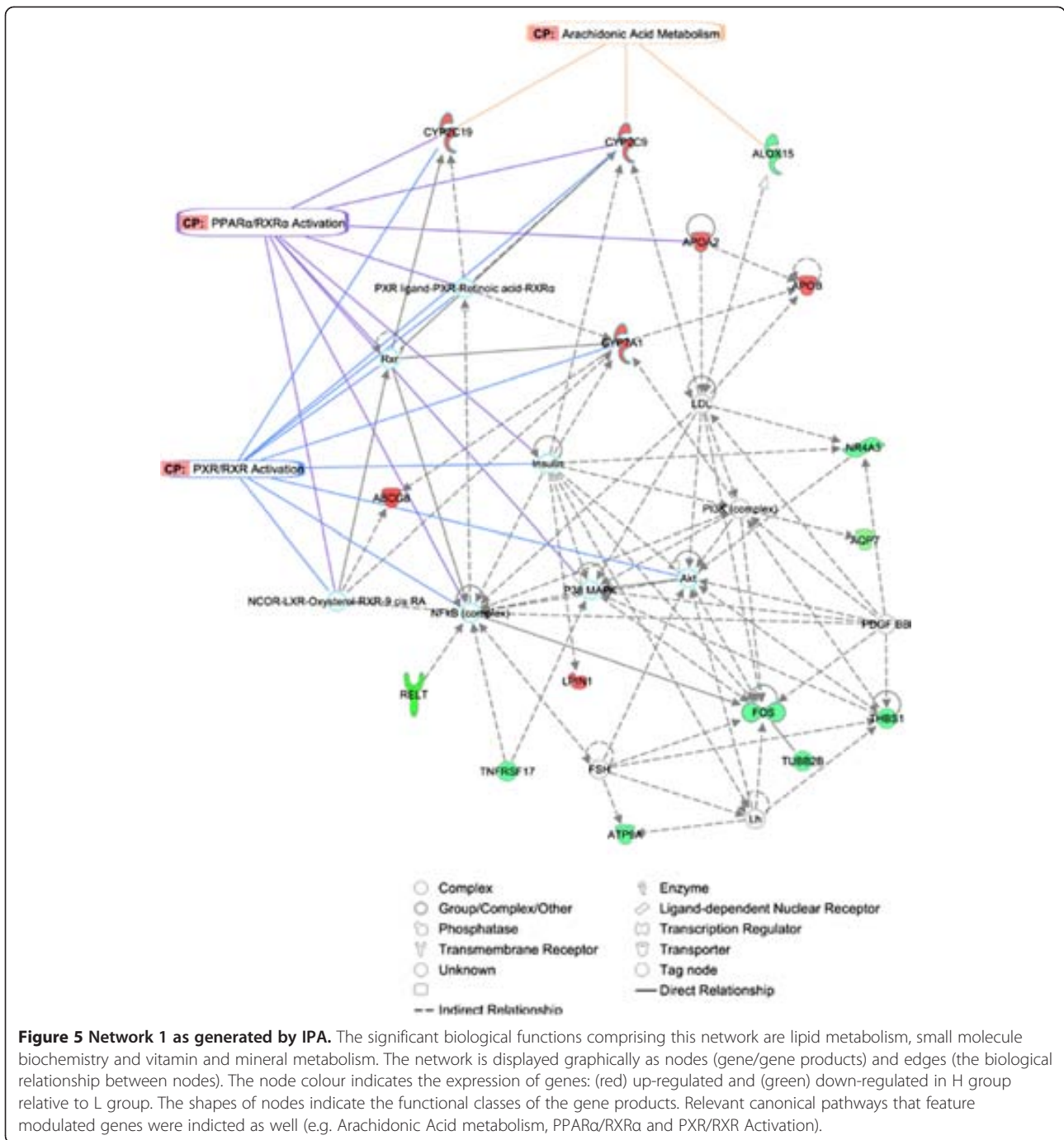
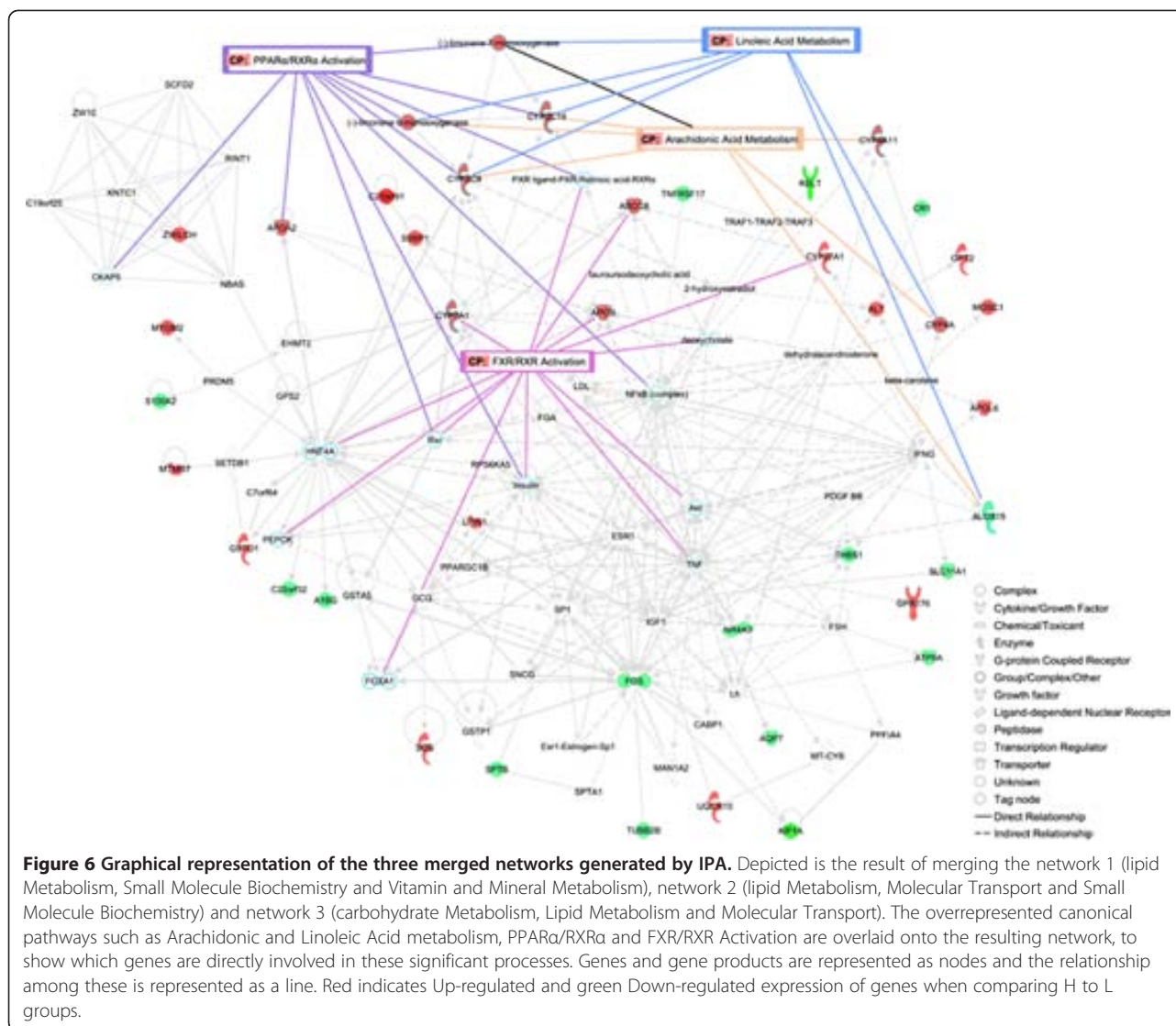


Figure 4 Canonical pathway significantly detected when contrasting the up (red) and (green) down-regulated genes in H compared to L group. X-axis values are the $\log(B-H \text{ correction } p\text{-value})$ and Y-axis values are the canonical pathways. The statistical significance of pathway modulation was calculated via a right-tailed Fisher's Exact test in Ingenuity Pathway analysis and represented as $-\log(P \text{ value})$: $-\log$ values exceeding 1.30 were significant false discovery rate (FDR) < 0.05.



There is increasing awareness of the wide range of health benefits of PUFA in general, and of ω -3 fatty acids in particular. Meat is an important basis of human nutrition, and pork meat is seen to be a major source of human food. The composition of fatty acids stored in adipose tissue in pigs largely reflects that of ingested lipids [5]. Thus, swine meat enriched with ω -3 fatty acids can be achieved by feeding with commercial diets supplemented with

this PUFA [8], and possibly by selective breeding. In fact, there is a genetic basis of PUFA level in pork meat. It is likely that H group presented higher absorption of essential PUFA, increasing their amount reaching the IMF tissue, which in turn could be considered as an important factor in the inhibition of the *de novo* saturated fatty acid proportion in meat. Furthermore, differences on elongation, desaturation and oxidation of those essential PUFA to longer-



chain ω -3 and ω -6 fatty acids cannot be discarded. Therefore, from the human health perspective, increasing H genotypes through breeding programs could be desirable because meat and meat-derived foods are still large contributors to saturated fatty acids intake in humans. These observations, together with several gene expression effects are the major factors leading us to believe that genetic was indeed a significant factor affecting meat IMF PUFA content and composition. However, an inverse relationship exists between nutritional value and eating quality of meat and, as consequence, established selection criteria to all together improve meat quality from the sensorial and nutritional point of view is a complex matter. Therefore, a holistic approach including both nutrigenetic and nutrigenomic disciplines may be required to improve the pork meat quality from both points of view.

Conclusions

We used RNA-Seq as a tool to explore the liver transcriptome of ten female pigs with extreme phenotypes for intramuscular fatty acid composition. Transposable elements, lncRNAs and new putative protein-coding genes were identified. Reproducibility of the data was confirmed by the strong correlation observed between the values of gene expression obtained by RNA-Seq, RT-qPCR and microarrays. A total of 55 genes differentially-expressed between extreme animals were identified. These genes belong to canonical pathways and gene networks related to the lipid and fatty acid metabolism. In concordance with the initial phenotypic classification, pathway analysis inferred that linolenic and arachidonic acid metabolism was altered between extreme animals. The results obtained may help in the design of new selection strategies to improve pork meat quality from both the sensorial and nutritional points of view.

Methods

Animal material and phenotypes

The population studied was originated by crossing three Iberian boars (Guadyrbas line) with 31 Landrace sows [55,56]. Five F1 males were backcrossed with 26 Landrace sows and 144 BC1_LD pigs were obtained. All pigs were raised in a normal intensive system, fed under standard management conditions and were slaughtered at an average age of 179.8 ± 2.6 days following national and institutional guidelines for the ethical use and treatment of animals in experiments.

A total of 48 traits related with growth, carcass quality and intramuscular fatty acid composition were measured. A PCA was performed with the *prcomp* procedure of R software [57], including phenotypic information from twenty-six of the total traits. Four of these were related to carcass quality (carcass height, weight of ham, weight of shoulder and intramuscular fat) whereas the rest corresponded to fatty acids composition in muscle and indices of fatty acids metabolism. Animals with extreme phenotypes, according to the first principal component, were selected to generate the High (H) and Low (L) groups with 20 animals per group (Figure 1). Phenotypic mean comparisons between groups were performed using R. Since sex differences in liver transcriptome have been reported in several species [58], selection was made considering pedigree information representing the parental genetic diversity and only females were retained for RNA sequencing (five per group).

RNA isolation, library preparation and sequencing

From the 10 selected animals, total RNA was isolated from liver using the *RiboPure™ Isolation of High Quality Total RNA* (Ambion®, Austin, TX) following the manufacturer's recommendations. RNA was quantified using the *NanoDrop ND-1000 spectrophotometer* (NanoDrop products, Wilmington, USA) and checked for purity and integrity in a *Bioanalyzer-2100* (Agilent Technologies, Inc., Santa Clara CA, USA).

Sequencing libraries were generated using *Illumina mRNA-Seq* following manufacturer's instructions and ten index codes were added to attribute sequences to each animal. A total of two channels of an *Illumina Hi-Seq 2000* instrument (*Fasteris SA*, Plan-les-Ouates, Switzerland) were used to sequence two pools of five samples (one pool with five samples per lane with barcoding).

Mapping, assembling and annotation of reads

After removal of sequencing adaptors and low-complexity reads *TopHat1.2.0* software [31] was employed to map reads using as reference the version 9.61 of pig genome (*Sscrofa* 9.61) [<http://www.ensembl.org/info/data/ftp/index.html>]. Quality control and reads statistics were determined with *FASTQC* [<http://www.bioinformatics>.

bbsrc.ac.uk/projects/fastqc/]. Transcript assembly was performed using *Cufflinks* v0.9.3 [59], with a minimum alignment count per *locus* of 10. Finally, *S-MART* [<http://urgi.versailles.inra.fr/Tools/S-MART>] for read annotation was used.

Gene expression quantification and correlation analysis with expression microarrays

Gene expression quantification was performed using the normalized number of fragments per kilobase of exon per million reads (FPKM) as reported in *Cufflinks* output [59]. Correlations between mean expression values between groups were calculated. All individuals were also assayed with high-density oligonucleotide microarray chips (*GeneChip® Porcine*) from *Affymetrix* (Santa Clara, CA) containing a total of 23,937 probe sets (23,256 transcripts), representing 20,201 *Sus scrofa* genes. Microarrays were hybridized and scanned at the *Institut de Recerca Hospital Universitari Vall d'Hebron* (Barcelona, Spain) following *Affymetrix* standard protocols. Expression data were generated with *Gene-Chip Operating Software (GCOS)*. Probes were adjusted for background noises and normalized using the *GCRMA* R package [60]. The average probe value per gene was calculated and a total of 6,025 Ensembl gene IDs could be retrieved to estimate the Spearman correlation between the log2 expression values of genes analysed by RNA-Seq and microarrays. Finally, a GO enrichment analysis with the *QuickGO* browser [<http://www.ebi.ac.uk/QuickGO/>] was performed for the top 100 most expressed genes.

Differential gene expression analysis

Differential expression analysis (DE) between groups was performed using *DESeq* [37]. This R package uses as input file the unambiguous table of counts per gene obtained from HTseq-count [<http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>]. *DESeq* models the data using negative binomial distributions assuming that the mean is a good predictor of variance. Therefore, it assumes that genes with similar expression level also have similar variance across replicates [37]. Following the *DESeq* author's recommendations, some exploratory diagnostic plots were executed to check the dispersion estimate and data quality. In order to ascertain the base variance the function 'varianceFitDiagnostics' was used and the per-gene estimates of the base variance was plotted against the base levels. The uniformity of the cumulative probabilities estimated by the 'varianceFitDiagnostics' was also verified via the 'residualsEcdfPlot' function.

Differentially-expressed genes were detected through the 'nbinomTest' function of *DESeq*. All the genes with a fold change between H and L groups higher than or equal to 1.5 fold were retained (total of 2051 genes). Then, for this subset of genes, the R package *q-value*

[61] was employed to calculate the false-discovery rate and genes with a p-value ≤ 0.005 (which is equivalent to a q-value ≤ 0.17) were retained.

Validation of differentially-expressed genes by RT-qPCR and copy number determination by qPCR

In order to evaluate the repeatability and reproducibility of gene expression data obtained by RNA-Seq, a RT-qPCR assay using SYBR Green chemistry (Fast Start Universal Sybr green master, Rox; Roche Applied Science, Mannheim, Germany) and the comparative Ct method [62] was performed in an ABI PRISM[®] 7900 HT (Applied Biosystems, Inc., FosterCity, CA).

The isolated RNA of individual samples was reverse-transcribed into cDNA using the *High Capacity Reverse cDNA transcription Kit* (Applied Biosystems) in a total volume of 20 μ l containing 1 μ g of total RNA, following the manufacturer's instructions. PCR primers were designed using Primer Express[™] software (Applied Biosystems) and are shown in Additional file 12, Table S4. Two genes: β -2 microglobulin (β 2m) and hypoxanthine phosphoribosyl-transferase 1 (HPRT1), previously validated as stable expressed control genes in liver tissue by geNorm [63] were used as endogenous controls (Corominas et al., unpublished data). Due to the comparative Ct method requiring the target and endogenous PCR efficiencies to be nearly equal, validation experiments for each gene were performed. Thus, the log cDNA dilution (1:2, 1:20, 1:200, 1:2,000) versus Δ Ct, was plotted to obtain absolute slopes < 0.1 in all cases that allowed the use of the $2^{-\Delta\Delta C_t}$ method. PCR amplifications were performed in a total volume of 20 μ l containing 5 μ l of cDNA sample diluted 1:125. Depending on the pair primers, various concentrations were utilized (see Additional file 12, Table S4). Each sample was analyzed by triplicate, thermal cycle was: 10 min at 95°C and 40 cycles of 15 sec at 95°C and 1 min at 60°C. A dissociation curve was drawn for each primer pair. Data was analyzed using the SDS v2.4 and DataAssist[™] v3.0 software (Applied Biosystems). The sample of lowest expression level was selected as calibrator. Correlation between RNA-seq (Htseq) and RT-qPCR data ($2^{-\Delta\Delta C_t}$) was carried out with R.

Copy number variation was quantified using the assay described above with some modifications. PCR amplification was carried out with 10 ng of genomic DNA isolated from diaphragm samples by the phenol-chloroform method [64]. Primers used to amplify *CYP2C49* gene are described in Additional file 12, Table S4. For single copy endogenous control gene amplification, a previously described design on the *glucagon* (*GCG*) gene [65] was used, but a single nucleotide substitution on primer forward was introduced to adapt the primer to the porcine species (Additional file 12, Table S4). A sample with the lowest copy number was selected as a calibrator.

Transposable element analysis

To identify repetitive and transposable elements in pig liver transcriptome RepeatMasker version open-3.3.0 [http://www.repeatmasker.org/] and the 'quick search' and 'pig' species options with Search Engine: NCBI/RMBLAST and complete Database: 20090604 were used.

Orthology and lncRNA detection

Intergenic expressed regions not yet annotated in the *Sscrofa* 9.61 pig genome assembly as described in [23] were extracted. Then, a conservative approach was followed, using only sequences expressed in at least four of the five animals of each group (H and L). To identify which of these transcripts were putative coding transcripts the Augustus software was used [36], providing exon boundaries and allowing complete protein translation from the forward strand. Finally, BLASTP was employed to check which of these predicted proteins were already annotated in the *Homo sapiens*, *Bos taurus* and *Sus scrofa* protein databases. For lncRNA annotation, the intergenic expressed regions were compared by BLAST with the 2,047 putative porcine lncRNA reported by Esteve-Codina et al. (2011). All transcripts that matched with an expectation value lower than 10^{-5} were retained.

Gene functional classification, network and canonical pathways analyses

Biological network generation, functional classification and pathways analyses of differentially-expressed genes were carried out using Ingenuity Pathways Analysis software (IPA; Ingenuity Systems, www.ingenuity.com). The list of human homologs that correspond to the 50 protein-coding pig genes differentially-expressed was uploaded into the application. Then, each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base (IPKB). Networks of these genes were generated based on their connectivity. Network analysis returns a score that ranks networks according to their degree of relevance to the network eligible molecules in the dataset [66]. The network score is based on the hypergeometric distribution and is calculated with the right-tailed Fisher's exact test. The score is the negative log of this p-value. Only those molecules that demonstrate direct and indirect relationships to other genes, or proteins were integrated into the analysis.

IPA Functional Analysis was employed to identify the most significant biological functions in the comparative dataset of H and L groups. A canonical pathways analysis was generated to identify the pathways from the IPA library that were most significant. Fischer's exact test was employed to calculate a p-value which determines the probability that each biological functions and/

or canonical pathway is due to chance alone. The cut-off for considering a significance association was established by Benjamini & Hochberg (B-H) multiple testing correction of the p-value (FDR < 0.05) [67].

Data availability

The full data sets have been submitted to Gene Expression Omnibus (GEO) under accession GSE38588 and at NCBI Sequence Read Archive (SRA) under Accession SRA053452, Bioproject: PRJNA168072.

Additional files

Additional file 1: Table S1. Phenotypic means comparison \pm standard deviation between the sequenced individuals.

Additional file 2: Figure S1. Profile of gene expression distribution in both High and Low groups.

Additional file 3: Figure S2. Correlations between expression values of genes analysed by both RNA-seq and Affymetrix microarray technologies. X-axis values are the log₂ of expression quantified with Affymetrix Microarray technology and y-axis are values of log₂ (FPKM).

Additional file 4: Table S2. Description of the repetitive elements identified in the pig liver transcriptome.

Additional file 5: Figure S3. Venn diagrams of the predicted lncRNA.

Additional file 6: Figure S4. Per-gene estimates of the base variance against the base levels. The red line represents the fit variance. X-axis values are the base mean and y-axis values are the log₁₀ of the base mean and y-axis values are the log₁₀ of the base variance.

Additional file 7: Figure S5. Curves of the empirical cumulative density functions in both H and L groups. X-axis values are the chi-squared probability of residual and y-axis values are the empirical cumulative density functions.

Additional file 8: Figure S6. Estimated variances as squared coefficients of variation produced with the function 'scvPlot'. X-axis values are the base mean and y-axis values are the squared coefficients of variation.

Additional file 9: Table S3. Comparison between RNA-seq (Htseq) and RT-qPCR (relative quantification) expression data of *APOA2*, *LPIN1*, *ME3*, *CYP7A1* and *CYP2C49* genes. Relative CNV data for *CYP2C49* in comparison to the reference individual H3 is indicated in the last column.

Additional file 10: Figure S7. Network 2 as generated by IPA. The significant biological functions comprising this network are Lipid Metabolism, Molecular Transport and Small Molecule Biochemistry. The network is displayed graphically as nodes (gene/gene products) and edges (the biological relationship between nodes). The node colour indicates the expression of genes: red up-regulated, green down-regulated in H group relative to L group. The shapes of nodes indicate the functional class of the gene product.

Additional file 11: Figure S8. Network 3 as generated by IPA. The significant biological functions comprising this network are Carbohydrate Metabolism, Lipid Metabolism and Molecular Transport. The network is displayed graphically as nodes (gene/gene products) and edges (the biological relationship between nodes). The node colour indicates the expression of genes: red up-regulated, green down-regulated in H group relative to L group. The shapes of nodes indicate the functional class of the gene product.

Additional file 12: Table S4. Primers designed for the validation of differentially-expressed genes by RT-qPCR and copy number determination by qPCR.

Abbreviations

IMF: Intramuscular fat; IPA: Ingenuity Pathways Analysis; ALA: α -linolenic acid; LA: Linolenic acid; EDA: Eicosadienoic acid; ETE: Eicosatrienoic acid; AA: Arachidonic acid; LPS: Endotoxin lipopolysaccharide; IL-1: Pro-

inflammatory cytokines; RXR: Retinoid X receptors; PXR: Pregnane X receptor; FXR: Farnesoid X receptor; PPAR- α : Peroxisome proliferator-activated receptors alpha.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YRC and JMF conceived and designed the experiment. JMF was the principal investigator of the project. YRC and AEC performed the RNA-Seq data analysis. YRC and NM performed the pathways analysis and drafted the manuscript. JE, AIF, MPE, NIE and JMF collected the samples. AC, MB and JC performed DNA and RNA isolation. AC, MB and JC performed the qPCR and RT-PCR assays. All authors read and approved the final manuscript.

Acknowledgements

This work was funded by MICINN projects AGL2008-04818-C03/GAN and AGL2011-29821-C02 (Ministerio de Economía y Competitividad), and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036, Centre for Research in Agrigenomics). Y. Ramayo-Caldas was funded by a FPU PhD grant from the Spanish Ministerio de Educación (AP2008-01450), J. Corominas was funded by a FPI PhD grant from the Spanish Ministerio de Educación (BES-2009-018223), A. Esteve-Codina is recipient of a FPI PhD fellowship from the Ministerio de Educación (BES-2008-005772), Spain. This manuscript has been proofread by Mrs. Valma Ruth Dunstan, BA, BEdSt, CELTA, MA (TESOL), a native English-speaker and instructor of English at the University of Queensland (Brisbane, Australia).

Author details

¹Centre de Recerca en Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra 08193, Spain. ²INRA, UMR 1313 Génétique Animale et Biologie Intégrative (GABI), Equipe Génétique Immunité Santé, Jouy-en-Josas F-78352, France. ³AgroParisTech, UMR 1313 GABI, Jouy-en-Josas F-78352, France. ⁴CEA, DSV/IRCM/SREIT/LREG, Jouy-en-Josas F-78352, France. ⁵Genètica i Millora Animal, IRTA Lleida, Lleida 25198, Spain. ⁶Departamento de Mejora Genética Animal, INIA, Ctra. De la Coruña km. 7, Madrid 28040, Spain. ⁷Institut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. ⁸Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain.

Received: 2 March 2012 Accepted: 8 October 2012

Published: 11 October 2012

References

- Cooper D: A brief history of cross-species organ transplantation. *Proc (Bayl Univ Med Cent)* 2012, **25**(1):49–57.
- Dodson M, Hausman G, Guan L, Du M, Rasmussen T, Poulos S, Mir P, Bergen W, Fernyhough M, McFarland D, Rhoads R, Soret B, Reedy J, Velleman S, Jiang Z: Lipid metabolism, adipocyte depot physiology and utilization of meat animals as experimental models for metabolic research. *Int J Biol Sci* 2010, **6**(7):691–699.
- Bauman DE, Mather IH, Wall RJ, Lock AL: Major Advances Associated with the Biosynthesis of Milk. *J Dairy Sci* 2006, **89**(4):1235–1243.
- FAO: *Fats and Fatty Acids in Human Nutrition. Report of Expert Consultation*; 2008:308. ISBN: 978-3-8055-9261-1.
- Wood JD, Richardson RI, Nute GR, Fisher AV, Campo MM, Kasapidou E, Sheard PR, Enser M: Effects of fatty acids on meat quality: a review. *Meat Sci* 2004, **66**(1):21–32.
- Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, Richardson RI, Hughes SJ, Whittington FM: Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci* 2008, **78**(4):343–358.
- Estévez M, Morcuende D, Cava López R: 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* 2003, **64**(4):499–506.
- Guillevic M, Kouba M, Mourot J: Effect of a linseed diet or a sunflower diet on performances, fatty acid composition, lipogenic enzyme activities and stearoyl-CoA-desaturase activity in the pig. *Livest Sci* 2009, **124**(1,Ái3):288–294.
- Pascual J, Rafecas M, Canela M, Boatella J, Bou R, Barroeta A, Codony R: Effect of increasing amounts of a linoleic-rich dietary fat on the fat

- composition of four pig breeds, Part II: Fatty acid composition in muscle and fat tissues. *Food Chem* 2007, **100**(4):1639–1648.
10. Casellas J, Noguera JL, Reixach J, Diaz I, Amills M, Quintanilla R: Bayes factor analyses of heritability for serum and muscle lipid traits in Duroc pigs. *J Anim Sci* 2010, **88**(7):2246–2254.
 11. Ntawubizi M, Colman E, Janssens S, Raes K, Buys N, De Smet S: Genetic parameters for intramuscular fatty acid composition and metabolism in pigs. *J Anim Sci* 2010, **88**:1286–1294.
 12. Gondret F, Ferre P, Dugail I: ADD-1/SREBP-1 is a major determinant of tissue differential lipidogenic capacity in mammalian and avian species. *J Lipid Res* 2001, **42**(1):106–113.
 13. Nafikov R, Beitz D: Carbohydrate and Lipid Metabolism in Farm Animals. *J Nutr* 2007, **137**(3):702–705.
 14. O'Hea E, Leveille G: Significance of adipose tissue and liver as sites of fatty acid synthesis in the pig and the efficiency of utilization of various substrates for lipogenesis. *J Nutr* 1969, **99**(3):338–344.
 15. Mourot J, Kouba M, Peiniau P: Comparative study of in vitro lipogenesis in various adipose tissues in the growing domestic pig (*Sus domesticus*). *Comp Biochem Physiol B Biochem Mol Biol* 1995, **111**(3):379–384.
 16. Nguyen P, Leray V, Diez M, Serisier S, Le Bloc'h J, Siliart B, Dumon H: Liver lipid metabolism. *J Anim Physiol Anim Nutr (Berl)* 2008, **92**(3):272–283.
 17. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 2008, **18**(9):1509–1517.
 18. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, **5**(7):621–628.
 19. Garber M, Grabherr MG, Guttman M, Trapnell C: Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Meth* 2011, **8**(6):469–477.
 20. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML: Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 2011, **12**(7):499–510.
 21. Cánovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano J: SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome* 2010, **21**(11):592–598.
 22. Huang W, Khatib H: Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. *BMC Genomics* 2011, **11**(1):711.
 23. Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C, Perez-Enciso M: Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. *BMC Genomics* 2011, **12**(1):552.
 24. Chen C, Ai H, Ren J, Li W, Li P, Qiao R, Ouyang J, Yang M, Ma J, Huang L: 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* 2011, **12**(1):448.
 25. Jager M, Ott C-E, Grunhagen J, Hecht J, Schell H, Mundlos S, Duda G, Robinson P, Lienau J: Composite transcriptome assembly of RNA-seq data in a sheep model for delayed bone healing. *BMC Genomics* 2011, **12**(1):158.
 26. Li R, Rinaldi M, Capuco A: Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. *Vet Res* 2011, **42**(1):114.
 27. Ramirez R, Cava R: Carcass composition and meat quality of three different Iberian x Duroc genotype pigs. *Meat Sci* 2007, **75**(3):388–396.
 28. Serra X, Gil F, Pérez-Enciso M, Oliver MA, Vázquez JM, Gispert M, Díaz I, Moreno F, Latorre R, Noguera JL: A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. *Livestock Production Science* 1998, **56**(3):215–223.
 29. Webb EC, O'Neill HA: The animal fat paradox and meat quality. *Meat Sci* 2008, **80**(1):28–36.
 30. Ramayo-Caldas Y, Mercadé A, Castelló A, Yang B, Rodríguez C, Alves E, Díaz I, Ibáñez-Escriche N, Noguera JL, Pérez-Enciso M, Fernández AI, Folch JM: Genome-Wide Association Study for Intramuscular Fatty Acid Composition in an Iberian x Landrace Cross. *J Animal Sci* 2012, **90**:1–11. doi:10.2527/jas2011-4900.
 31. Trapnell C, Pachter L, Salzberg SL: TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009, **25**(9):1105–1111.
 32. Zytynski M, Quesneville H: S-MART, A Software Toolbox to Aid RNA-seq Data Analysis. *PLoS One* 2011, **6**(10):e25988.
 33. Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009, **10**(1):57–63.
 34. Yu Y, Ping J, Chen H, Jiao L, Zheng S, Han Z-G, Hao P, Huang J: A comparative analysis of liver transcriptome suggests divergent liver function among human, mouse and rat. *Genomics* 2010, **96**(5):281–289.
 35. Ramskold D, Wang ET, Burge CB, Sandberg R: An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. *PLoS Comput Biol* 2009, **5**(12):e1000598.
 36. Stanke M, Diekhans M, Baertsch R, Haussler D: Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 2008, **24**(5):637–644.
 37. Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biol* 2010, **11**(10):R106.
 38. Kirischian N, McArthur A, Jesuthasan C, Krattenmacher B, Wilson J: Phylogenetic and Functional Analysis of the Vertebrate Cytochrome P450 2 Family. *J Mol Evol* 2011, **72**(1):56–71. doi:10.1007/s00239-010-9402-7.
 39. Ramayo-Caldas Y, Castello A, Pena R, Alves E, Mercade A, Souza C, Fernandez A, Perez-Enciso M, Folch J: Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. *BMC Genomics* 2010, **11**(1):593.
 40. Wise EM Jr, Ball EG: Malic Enzyme and Lipogenesis. *Proc Natl Acad Sci U S A* 1964, **52**:1255–1263.
 41. Cherkaoui-Malki M, Meyer K, Cao WQ, Latruffe N, Yeldandi AV, Rao MS, Bradford CA, Reddy JK: Identification of novel peroxisome proliferator-activated receptor alpha (PPARalpha) target genes in mouse liver using cDNA microarray analysis. *Gene Expr* 2001, **9**(6):291–304.
 42. Tugwood JD, Holden PR, James NH, Prince RA, Roberts RA: A peroxisome proliferator-activated receptor-alpha (PPARalpha) cDNA cloned from guinea-pig liver encodes a protein with similar properties to the mouse PPARalpha: implications for species differences in responses to peroxisome proliferators. *Arch Toxicol* 1998, **72**(3):169–177.
 43. Cheon Y, Nara TY, Band MR, Beever JE, Wallig MA, Nakamura MT: Induction of overlapping genes by fasting and a peroxisome proliferator in pigs: evidence of functional PPARC± in nonproliferating species. *Am J Physiol Regul Integr Comp Physiol* 2005, **288**(6):R1525–R1535.
 44. Finck BN, Gropler MC, Chen Z, Leone TC, Croce MA, Harris TE, Lawrence JC Jr, Kelly DP: Lipin 1 is an inducible amplifier of the hepatic PGC-1C±/PPARC± regulatory pathway. *Cell Metab* 2006, **4**(3):199–210.
 45. Barnes KM, Winslow NR, Shelton AG, Hlusko KC, Azain MJ: Effect of dietary conjugated linoleic acid on marbling and intramuscular adipocytes in pork. *J Anim Sci* 2011, **90**(4):1142–1149.
 46. Theil PK, Lauridsen C: Interactions between dietary fatty acids and hepatic gene expression in livers of pigs during the weaning period. *Livest Sci* 2007, **108**(1):26–29.
 47. Pascual J, Rafecas M, Canela M, Boatella J, Bou R, Baucells M, Codony R: Effect of increasing amounts of a linoleic-rich dietary fat on the fat composition of four pig breeds, Part I: Backfat fatty acid evolution. *Food Chem* 2006, **96**(4):538–548.
 48. Chizzolini R, Novelli E, Zanardi E: Oxidation in traditional mediterranean meat products. *Meat Sci* 1998, **49**(Supplement 1(0)):S87–S99.
 49. Carrapiso AI, Bonilla F: Garcf±ÁÁa C: Effect of crossbreeding and rearing system on sensory characteristics of Iberian ham. *Meat Sci* 2003, **65**(1):623–629.
 50. Katan M, Zock P, Mensink R: Effects of fats and fatty acids on blood lipids in humans: an overview. *Am J Clin Nutr* 1994, **60**(6):1017S–1022S.
 51. Chizzolini R, Zanardi E, Dorigoni V, Ghidini S: Calorific value and cholesterol content of normal and low-fat meat and meat products. *Trends in Food Science & Technology* 1999, **10**(4–5):119–128.
 52. Rudel LL, Parks JS, Sawyer JK: Compared With Dietary Monounsaturated and Saturated Fat, Polyunsaturated Fat Protects African Green Monkeys From Coronary Artery Atherosclerosis. *Arterioscler Thromb Vasc Biol* 1995, **15**(12):2101–2110.
 53. Poudyal H, Panchal SK, Diwan V, Brown L: Omega-3 fatty acids and metabolic syndrome: Effects and emerging mechanisms of action. *Prog Lipid Res* 2011, **50**(4):372–387.
 54. Stables MJ, Gilroy DW: Old and new generation lipid mediators in acute inflammation and resolution. *Prog Lipid Res* 2011, **50**(1):35–51.
 55. Clop A, Ovilo C, Perez-Enciso M, Cercos A, Tomas A, Fernandez A, Coll A, Folch J, Barragan C, Diaz I, Oliver M, Varona L, Sillio L, Sanchez A, Noguera J: Detection of QTL affecting fatty acid composition in the pig. *Mamm Genome* 2003, **14**(9):650–656.
 56. Perez-Enciso M, Clop A, Noguera JL, Ovilo C, Coll A, Folch JM, Babot D, Estany J, Oliver MA, Diaz I, Sanchez A: A QTL on pig chromosome 4 affects

- fatty acid metabolism: evidence from an Iberian by Landrace intercross. *J Anim Sci* 2000, **78**(10):2525–2531.
57. Ihaka R, Gentleman R: **R: A Language for Data Analysis and Graphics.** *Journal of Computational and Graphical Statistics* 1996, **5**(3):299–314.
 58. Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, Waxman DJ: **Transcriptional Profiling of Human Liver Identifies Sex-Biased Genes Associated with Polygenic Dyslipidemia and Coronary Artery Disease.** *PLoS One* 2011, **6**(8):e23506.
 59. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** *Nat Biotech* 2010, **28**(5):511–515.
 60. Zhijin W, Irizarry RA, Robert G, Francisco Martinez M, Spencer F: *A Model Based Background Adjustment for Oligonucleotide Expression Arrays.* Johns Hopkins University; Dept. of Biostatistics; 2004. Available in: <http://www.bepress.com/jhubiostat/paper1>.
 61. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100**(16):9440–9445.
 62. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative CT method.** *Nat Protocols* 2008, **3**(6):1101–1108.
 63. Overgard A-C, 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**(1):36. doi:10.1186/1471-2199-11-36.
 64. Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K: *Current Protocols in Molecular Biology*; 1987.
 65. Ballester M, Castelló A, Ibáñez E, Sánchez A, Folch J: **Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals.** *Biotechniques* 2004, **37**(4):610–613.
 66. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF: **Large Scale Collab. Res P, Host Response to I: A network-based analysis of systemic inflammation in humans.** *Nature* 2005, **437**(7061):1032–1037.
 67. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995, **57**(1):289–300.

doi:10.1186/1471-2164-13-547

Cite this article as: Ramayo-Caldas et al.: Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics* 2012 **13**:547.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



CHAPTER VI

ANNEXES

From SNP co-association to gene co-expression:

Evidence of pleiotropic expression-QTL in lipid metabolism in pigs

Authors: Y. Ramayo-Caldas^{1,2}, M. Ballester^{1,2}, M. R. S. Fortes³, A. Esteve-Codina^{1,2}, A. Castelló², J.L. Noguera⁴, A.I. Fernández⁵, M. Pérez-Enciso^{1,6}, A. Reverter⁷, J. M. Folch.^{1,2}

Affiliations: ¹Centre de Recerca en Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, 08193, Spain. ²Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain. ³The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Center for Animal Science, Gatton, Queensland 4343, Australia. ⁴IRTA, Genètica i Millora Animal. 25198 Lleida, Spain. ⁵INIA, Mejora Genética Animal, 28040 Madrid. ⁶Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Spain. ⁷Commonwealth Scientific and Industrial Research Organisation, Division of Animal, Food and Health Sciences, Brisbane, Queensland 4067, Australia.

Abstract

Fatty acids (FA) play a critical role in energy homeostasis and metabolic diseases; in the context of livestock species, their profile also impacts on meat quality for healthy human consumption. In this study, genome-wide association results across 15 phenotypes were subjected to an association weight matrix (AWM) approach to predict a network of 1,096 genes related to intramuscular FA composition in pigs. To identify the key regulators of FA metabolism, we focused on the minimal set of transcription factors (TF) that explored majority of the network topology. This search resulted in a trio of TF comprised of *Nuclear receptor coactivator 2 (NCOA2)*, *E1A binding protein p300 (EP300)* and *four and a half LIM domains 2 (FHL2)*. Promoter sequence analyses confirmed that these TF have binding sites for some well-know regulators of lipid and carbohydrate metabolism. For the first time in a vertebrate, some of the co-associations observed at the genetic (SNP) level were validated through co-expression at the transcriptomic level based on real-time PCR of 40 genes in adipose tissue and a further 55 genes in liver. In particular, liver expression of *NCOA2* and *EP300* differed between pig breeds (Iberian and Landrace) extreme in terms of fat deposition. Highly clustered co-expression networks in both liver and adipose tissues were observed. *EP300* and *NCOA2* showed centrality parameters above average in the both networks. Over all genes, co-expression analyses confirmed 28.9% of the AWM predicted gene-gene interactions in liver and 33.0% in adipose tissue. The magnitude of this validation varied across genes, with up to 60.8% of the connections of *NCOA2* in adipose tissue being validated via co-expression. Our results recapitulate the known transcriptional regulation of FA, predict gene interactions that can be experimentally validated, and provides additional evidence supporting the use of pigs as an animal model for human metabolic diseases.

Introduction

Pigs (*Sus scrofa*) are originally from Southeast Asia, but today the domestic pig is common worldwide (Groenen et al., 2012; Larson et al., 2005). This global presence is evidence of pig relevance to humans, both as a source of food and as an animal model for scientific progress. Studies on the genetic basis of economically important traits in pigs are of great interest for animal production (Clop et al., 2003; Esteve et al., 2011; Fernandez et al., 2012; Hernandez-Sanchez et al., 2012; Mote et al., 2009; Perez-Enciso et al., 2000; Ramayo-Caldas et al., 2012b). In medical research, pigs are important for the development of xenotransplantation technology and as models for metabolic diseases such as obesity, type II diabetes (T2D) and atherosclerosis (Bendixen et al., 2010; Ekser et al., 2012; Houpt et al., 1979; Seki et al., 2012). Technological, nutritional and organoleptic properties of pork meat quality are highly dependent on lipid content and fatty acid (FA) composition (Wood et al., 1997; Wood et al., 1999; Wood et al., 2008). Further, FA are a major energy source and important constituents of cell membranes, playing a relevant role as cellular signaling molecules in various metabolic pathways, including metabolic diseases (Wakil et al., 2009). Environmental and genetic effects determining FA composition in pigs have been the subject of many studies. Supporting a genetic influence on FA composition moderate to high heritabilities have been reported (Casellas et al., 2010; Ntawubizi et al., 2010). However, the molecular process controlling FA composition and metabolism is far from being fully understood. Elucidating this molecular process could aid technological development towards improvement of pork meat quality and increased knowledge of FA metabolism, underpinning metabolic diseases.

Molecular pathways controlling lipid metabolism are highly interconnected. Also, they interact with other related pathways, such as carbohydrate metabolism and energy homeostasis pathways. Together, these pathways and its interactions constitute an essential metabolic network for homeostatic control and normal organism development (Hardie, 2012). In this context, a system biology approach focused on the connections and functional interactions between genes that underpin these metabolic pathways is an attractive alternative to the classical “single-gene – single-trait” approach found in most genome-wide association studies (GWAS) using single nucleotide polymorphisms (SNP).

The main goal of this study was to employ a previously described system biology approach termed Association Weight Matrix (AWM) (Fortes et al., 2010) and, based on a SNP-to-SNP co-association evidence, infer a gene network for intramuscular (IMF) FA composition in pigs. This multi-trait approach was applied to data from 15 phenotypes related to FA composition and metabolism. The analysis of the predicted gene network revealed key transcription factors that are network hubs and would be critical to determining meat quality, FA composition and controlling energy homeostasis in pigs. Finally, we experimentally validated some of the AWM network predictions using real-time PCR gene co-expression analyses in adipose and liver tissues.

Results

Genotyping data from 48,119 SNPs in 144 backcross pigs (25% Iberian × 75% Landrace) was employed for GWAS of fatty acid related traits in the *Longissimus dorsi* muscle. Manhattan plots representing GWAS results for each of 15 phenotypes analyzed traits are provided in Supplementary Figure SF1. For all 15 phenotypes, estimated SNP additive effects were standardized (z-scores) by subtracting the mean and dividing by the

phenotype-specific standard deviation. After applying a series of selection criteria (see Methods), a total of 1,096 SNPs were retained to build the AWM matrix. Correlations between phenotypes were calculated using AWM columns (standardized SNP effects across traits) and were visualized as a hierarchical tree cluster, in which strong positive and negative correlations are displayed as proximity and distance, respectively (Figure 5). As expected, FA with physiological similarities, such as palmitic acid and saturated FA (SFA), oleic with monounsaturated FA (MUFA) and linoleic with polyunsaturated FA (PUFA), cluster together (Figure 5).

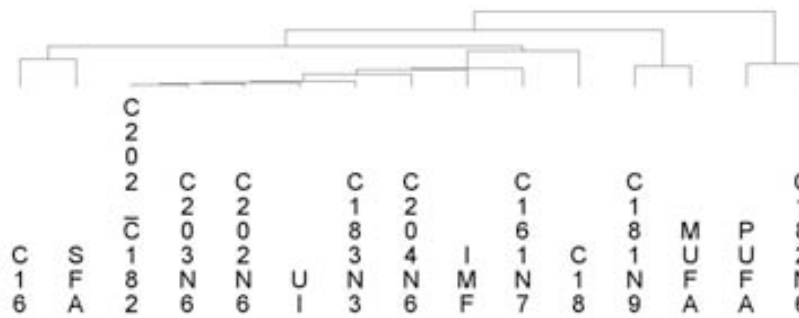


Figure 5. Hierarchical cluster analysis of the 15 phenotypes analyzed in this study: Palmitic acid (C16), Stearic acid (C18), Palmitoleic acid (C16N7), Oleic acid (C18N9), Linoleic acid (C18N6), α -Linolenic acid (C18N3), Eicosadienoic acid (C20N6), Eicosatrienoic acid (C20N6), Arachidonic acid (C20N6), Saturated FA (SFA), Monounsaturated FA (MUFA), Polyunsaturated FA (PUFA), Unsaturated indices (UI), Elongase activity (C20|C18), Percentage intramuscular fat (IMF).

Gene interactions were predicted using pair-wise correlation analysis of the SNP effects across pair-wise rows of the AWM. Hence, the AWM predicted gene interactions based on significant co-association between SNPs. In the network, every node represents a gene (or SNP), whereas every edge connecting two nodes represents a significant interaction. In total, 111,198 significant edges (or 18.5% of all the possible edges) between the 1,096 nodes were identified as significant by the PCIT algorithm (Reverter et al., 2008) (Figure 6A). For every node we computed the total number of connections based on significant interactions. Table 6 list the ten most connected nodes and Supplementary Table ST1 their positional concordance with fat-related QTL deposited in the Pig QTL Database.

Table 6. Description of the ten most connected nodes in the co-association network.

SNP/Gene	Illumina Chip SNP	Associated Traits	Connections	Sequence
ALGA0061664	ALGA0061664	3	376	Intergenic variant
SLC30A9	H3GA0024739	1	373	Intronic variant
SEMA3F	DIAS0001129	3	370	Intronic variant
ARHGEF2	ASGA0021047	3	368	Downstream variant
NTRK3	MARC0045253	3	367	Intronic variant
ZFHX4	ALGA0025325	1	366	Intronic variant
SLC22A3	ALGA0117149	6	365	Intronic variant
ARMC4	ALGA0059185	4	356	Intronic variant
C9orf171	ASGA0008154	6	355	Intronic variant
PPP2R2A	DIAS0004697	5	353	Splice region variant

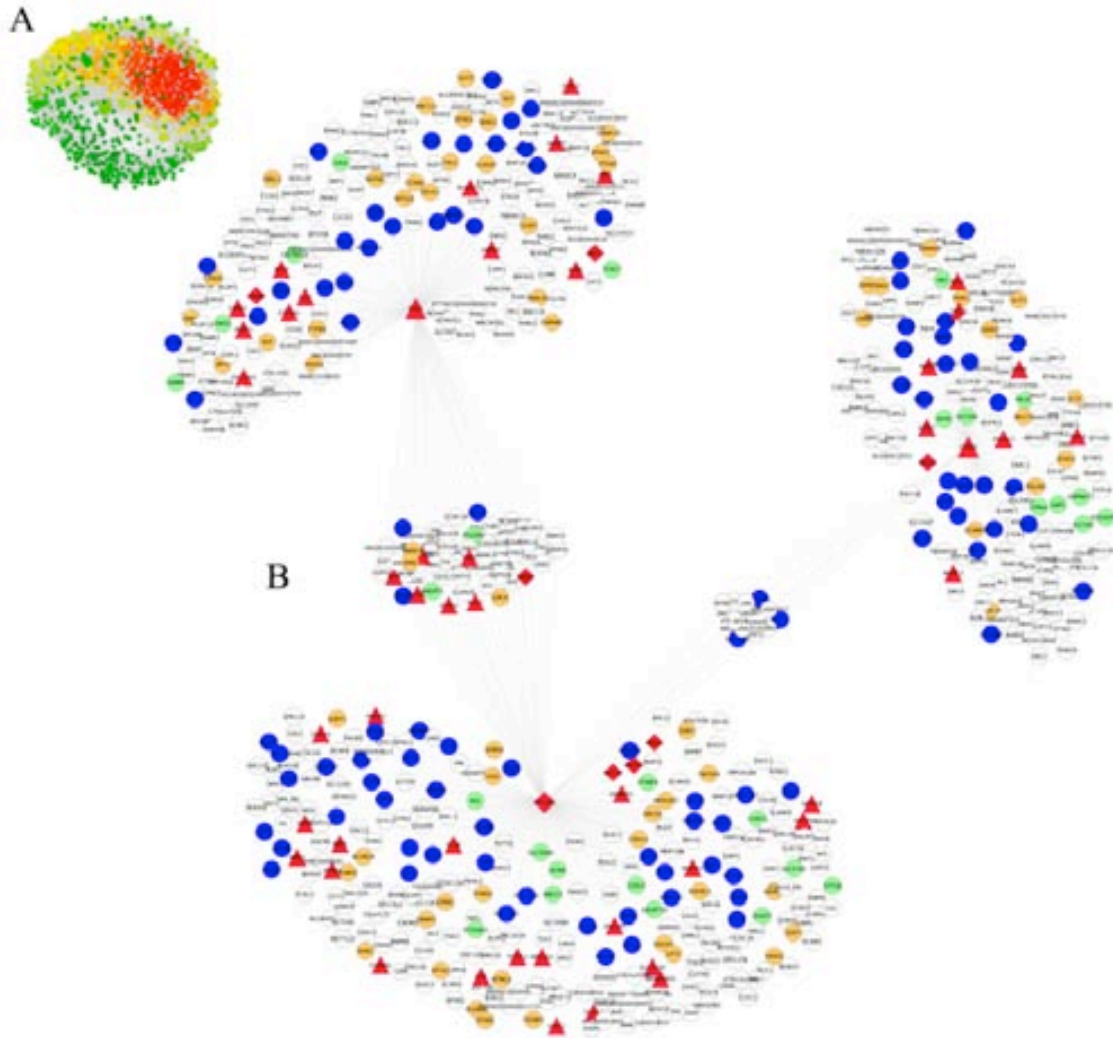


Figure 6. Co-association network based on the AWM approach. (A) Entire network with 1,096 nodes (i.e., genes or SNPs) and 111,198 interactions. The color spectrum ranges from green to red for low and high density, respectively. (B) Subset of the network showing the best trio of transcription factors: *NCOA2*, *EP300* and *FHL2*. Node color corresponding with the functional classification of the in silico predicted target gene as follows: TF (red), lipid metabolism process (blue), carbohydrate metabolisms (green), development process (orange) and finally, white nodes represent genes with others functional classification. Node shape indicates classification as: diamond (TF involved in lipid metabolism), triangle (TF), ellipse (other genes).

Gene ontology (GO) and pathway enrichment analyses were performed to gain insight into the predicted gene network. Overrepresented GO terms in the network included: “Cellular component organization” ($P = 4.74 \times 10^{-10}$, FDR = 1.82×10^{-8}), “Anatomical structure morphogenesis” ($P = 3.49 \times 10^{-6}$, FDR = 3.88×10^{-5}), “Kinase activity” ($P = 1.02 \times 10^{-5}$, FDR = 9.04×10^{-5}), “Hydrolase activity” ($P = 5.66 \times 10^{-4}$, FDR = 2.05×10^{-3}), “Lipid metabolic process” ($P = 5.71 \times 10^{-4}$, FDR = 2.05×10^{-3}), “Enzyme regulator activity” ($P = 2.21 \times 10^{-3}$, FDR = 3.78×10^{-3}) and “Lipid binding” ($P = 1.34 \times 10^{-2}$, FDR = 3.06×10^{-2}). Supplementary Table ST2 provides the full list of overrepresented GO terms. A detailed examination of the genes underpinning the enrichment for “Lipid metabolic process” revealed that most of these belong to three functional categories: “Metabolism of membrane lipid derivative” (32 genes, $P = 9.67 \times 10^{-6}$), “Synthesis of lipid” (45 genes, $P = 3.69 \times 10^{-4}$) and “Fatty acid metabolism” (36 genes, $P = 1.39 \times 10^{-3}$). Pathway analyses revealed an enrichment for “Regulation of actin cytoskeleton (hsa04810)”, “Focal adhesion (hsa04510)”, “Pathways in cancer (hsa05200)”, “Chemokine signalling pathway (hsa04062)”, “Phosphatidylinositol signalling system (hsa04070)” and “Inositol phosphate metabolism (hsa00562)” (Supplementary Table ST3).

To identify potential regulators of the above-mentioned pathways and GO categories, we focussed on TF found in the gene network. We applied an information lossless approach that explored the 64,824 possible trios among the available 74 TF (see Methods and Supplementary Table ST4 for complete list of TF) and identified the TF trio that spanned most of the network topology with minimum redundancy. These three TF were: *Nuclear receptor coactivator 2 (NCOA2, alias TIF2)*, *E1A binding protein p300*

(*EP300*, alias *p300*) and *four and a half LIM domains 2* (*FHL2*, alias *SLIM-3*). Interestingly, these TF contain transcription factors binding sites (TFBS) in their promoter region for some well know TF that are considered as important regulators of lipid and carbohydrate metabolism such as: *SREBP-1*, *PPARG*, *PPAR- α* , *HNFI1A*, *HNFI4- α* , *ER- α* and *GR- α* . In the predicted network, a total of 730 genes show co-association with the three key TF (Figure 6 B). A detailed examination of the most representative pathways related to these 730 predicted target genes showed a significant overrepresentation for “HIF-1 signaling pathway (hsa04066)”, “Acute myeloid leukemia (hsa05221)”, “Colorectal cancer (hsa05210)”, “Renal cell carcinoma (hsa05211)” and “Type II diabetes mellitus (hsa04930)” (Supplementary Figure SF2). Admittedly, some of the above-mentioned GO terms and pathways could have been expected from a network predicted from GWAS of FA-related phenotypes and this gives confidence in the reliability of the results. Others, however, were unexpected and might lead to new insights on FA physiology.

Functional validation: From co-association to co-expression analysis in liver and adipose tissues

The expression of the three TF across *Longissimus dorsi* muscle (LD), adipose and liver tissue was explored. In concordance with previous results, suggesting that highly connected TF are in general broadly expressed across tissues (Ravasi et al., 2010), the three TF were expressed across all the studied tissues. Further, a comparison between Iberian and Landrace pig breeds revealed significant increase fold changes (FC) in the liver of Iberian pigs for the expression of *NCOA2* (FC = 1.56, $P < 0.01$) and *EP300* (FC = 1.23, $P < 0.05$) (Figure 7).

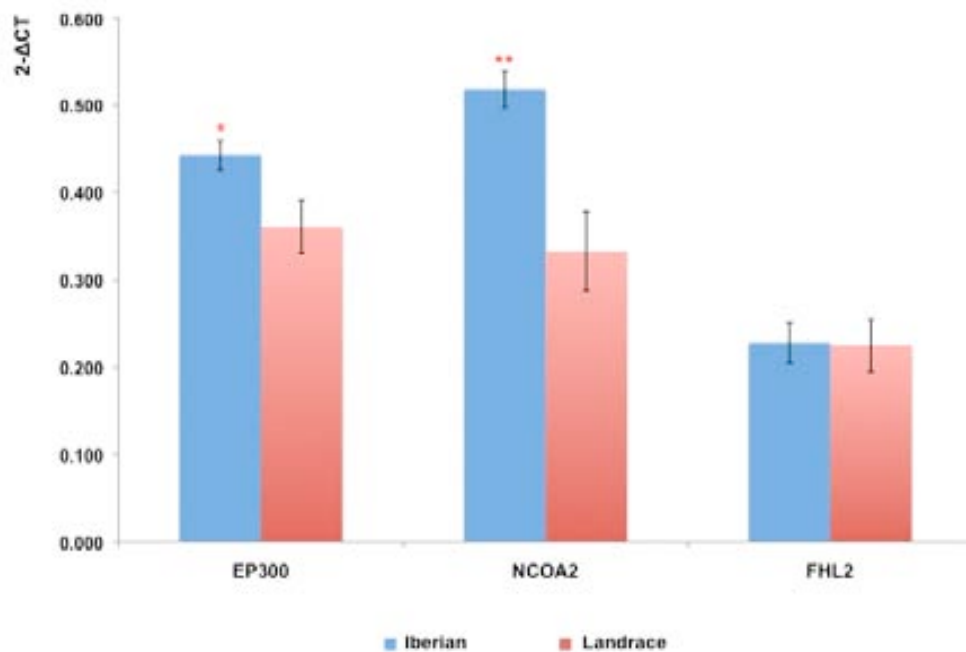


Figure 7. Results of the liver differential expression analysis comparing the best TF trio in the Iberian and Landrace breeds.

The expression patterns of 43 genes in liver and 40 genes in adipose tissue were successfully measured across 55 backcross animals. In liver, the expression data of twelve additional genes were also included in the co-expression analysis (see Methods). Co-expression analysis revealed highly connected networks in both liver and adipose tissue, suggesting strong functional interconnections among the studied genes. Topology of liver co-expression network showed 55 nodes connected by 425 edges (Supplementary Figure SF3A) and in adipose tissue 40 nodes and 261 edges were observed (Supplementary Figure SF3B). Network parameters such as average degree (Deg) and average distance (AvD_G) were slightly higher in liver co-expression network compared to adipose tissue network ($Deg_{Liver} = 15.45$ $AvD_G = 1.81$ vs $Deg_{Adipose} = 13.05$ and $AvD_G =$

1.75). Based on network centrality, the relevance of individual genes differs within each network. For example, topological properties of the liver co-expression network suggest an important role for *ARNT* in the regulation of hepatic lipogenic and gluconeogenesis activity, and these findings agree with published results (Rankin et al., 2009a; Wang et al., 2009). It should be noted that *BCL9* showed the highest centrality value in the liver co-expression network (Supplementary Figure SF3A). In addition, degree analysis showed that *BCL9*, *EP300*, *PBX1*, *SIRT1*, *PIP5K1A* and *ARNT* were the most central genes in the liver co-expression network. However, in the adipose co-expression network, degree analysis suggested that *ANK2*, *NCOA2*, *SIRT1*, *EIF4E*, *HMBOX1* are the most central genes (Supplementary Figure SF3B). When analysing a sub-network of the liver co-expression network, formed only by the same 40 genes included in the adipose co-expression network, five genes (*BCL9*, *EP300*, *PBX1*, *PIP5K1A*, and *SIRT1*) were still the most central genes and this finding underscores their relevant role in the function and structure of the liver co-expression network.

Beyond the study of the topological properties of the liver and adipose tissue co-expression networks, we were concerned with those, if any gene interactions predicted via SNP co-association were corroborated through co-expression analyses. In line with recent results in yeast (Wang et al., 2013), we observed that interacting loci could jointly regulate the co-expression patterns of pairs of genes. For the first time, in a not model species co-expression analyses confirmed AWM predicted gene-gene interactions based on SNP co-association. However, it is worth noting that the magnitude of this validation varied in a tissue-specific manner. For instance, with respect to the liver module formed by 48 AWM nodes and 359 edges (based on co-association analysis) we observed that 28.9% (104/359) of the predicted gene-gene interactions were validated by the co-

expression results (Figure 8A). Whereas in the adipose tissue, the observed percentage of the AWM validated interactions was slightly higher representing 33.0% of the possible combinations (Figure 8B). When we limited this comparison to the intersecting 39 genes included in both co-expression networks, the proportion of the AWM gene-gene interactions validated in liver (29.5%) was still lower than in adipose tissue (33.0%). Comparing both networks, we observed that approximately 35.7% (or 30 out of 84) of the interactions validated in the adipose tissue were also validated in the liver co-expression analysis (Supplementary Table ST5). Interestingly, these always co-associated and co-expressed genes belong to biological processes related to lipid metabolism including: Negative Regulation of Fat Cell Differentiation (*INSIG1, TCF7L2, ZFPM2*), Androgen Receptor Signalling Pathway (*EP300, FHL2, NCOA2*), Response to Hormone Stimulus (*ABCC5, ANGPT1, FABP3, EP300, SORT1, FHL2*) and Lipid Metabolic Process (*PBX1, INSIG1, FABP3, FDFT1, PIP5K1A, MAX, AASDH*). When we focused on the best TF trio, we observed that 60.8% (or 14 out of 23) of the interactions of *NCOA2* predicted by the AWM co-association network were corroborated in the co-expression network of the adipose tissue. This percentage dropped to 34.6% (or 9 out of 26) in the co-expression network of the liver tissue. For *EP300*, 44.4% (or 4 out of 9) of the AWM predicted interactions were observed in the adipose co-expression network and 41.6% (5 out of 12) in the liver co-expression network. Finally, for *FHL2* we observed the lowest percentage of validated interactions: 20.0% (or 2 out of 10) in adipose tissue and 14.3% (2 out of 14) in liver (Table 7).

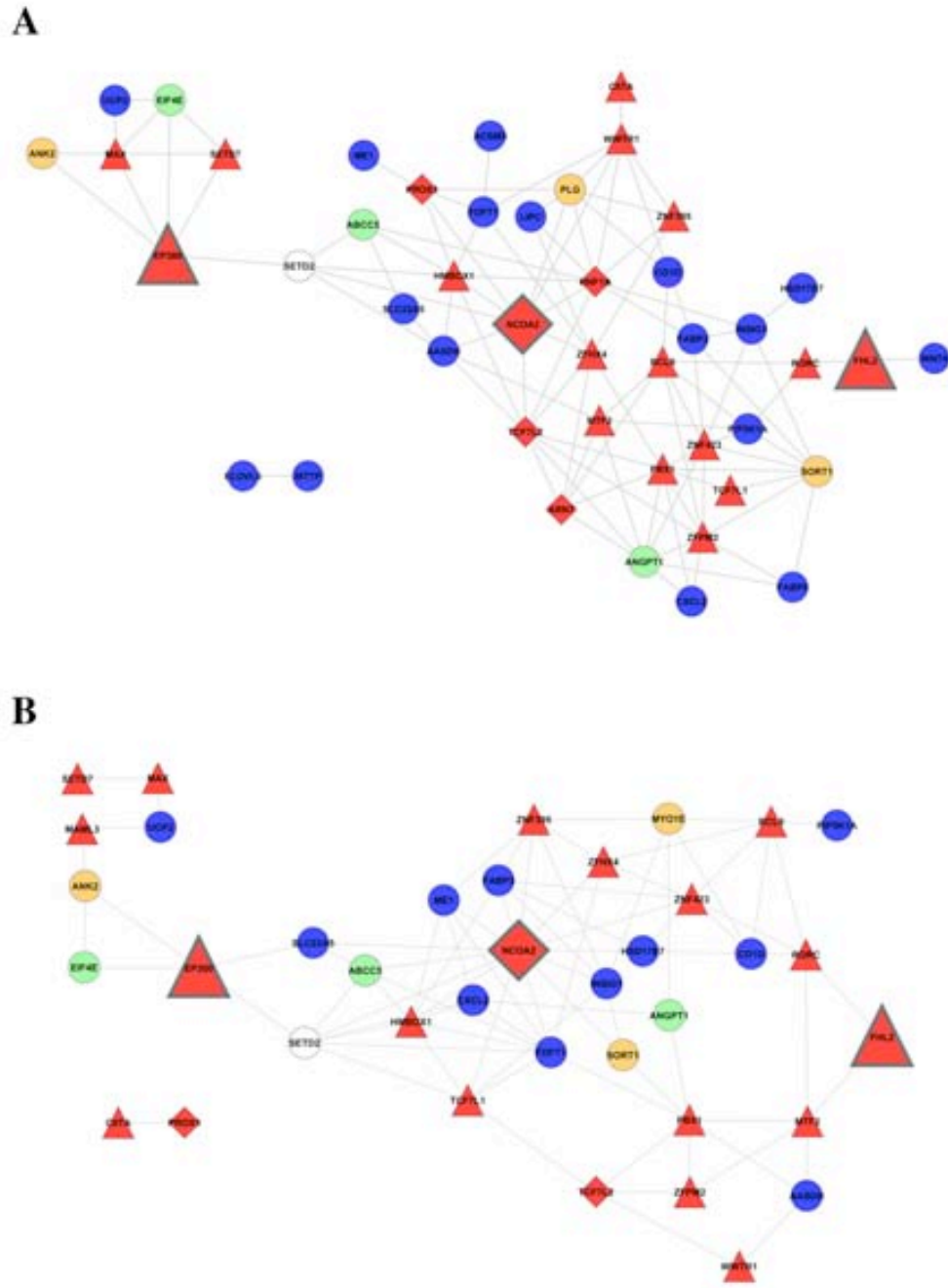


Figure 8. Connections from the co-association network that were confirmed by the co-expression network in liver (A) and adipose (B) tissue. Nodes color relate to the functional classification of genes as follows: TF (red nodes), lipid metabolism (blue

nodes), carbohydrate metabolism (green), development process (orange) and white nodes represent genes with others functional classification. The size of the nodes corresponding to the best trio of transcription factors (*NCOA2*, *EP300* and *FHL2*) has been enlarged to facilitate their location.

Table 7. Concordance validation between the co-association and the co-expression networks for the best TF trio and in the two tissues.

Tissue	TF	Connections in the AWM co-association network ^A	Connections in the qPCR co-expression network ^A	Validated connections	% Validation
Adipose	NCOA2	23	21	14	60.8
	EP300	9	19	4	44.4
	FHL2	10	9	2	20.0
Liver	NCOA2	26	18	9	34.6
	EP300	12	28	5	41.6
	FHL2	14	13	2	14.3

^A Connections deemed significant according to the PCIT algorithm.

Discussion

Molecular processes controlling FA metabolism are highly interconnected and linked with related pathways, such as lipid, carbohydrate and energy metabolism. In fact, FA are a major energy source and together with several factors, such as total energy intake, dietary fat/carbohydrate ratio, or glucose and/or insulin concentration, regulate *de novo* lipogenesis (Parks et al., 1999; Picard et al., 2002). As a consequence, it is expected that at the selected threshold ($P < 0.035$) our best trio of TF (*NCOA2*, *EP300*, *FHL2*) show co-association with an large number of genes and other TF relevant for lipid, carbohydrate and energy metabolism. For instance, 39 of the predicted target genes via

SNP co-association (Supplementary Table ST6) have been recently reported in two large-scale meta-analysis studies for plasma lipids in humans (Asselbergs et al., 2012; Teslovich et al., 2010). Interestingly, many of these genes, including our TF trio and other FA relevant genes, would have been missed by traditional single-trait GWAS due to the lack of an acceptably significant association level (i.e. $P > 0.05$ after correction for multiple testing). As noted before (Fortes et al., 2010) and confirmed by this study, AWM points to new candidate genes, TF and gene interactions via exploring SNP co-associations across multiple traits beyond the one-dimensional approach for identifying genes affecting single traits.

In the predicted network, *NCOA2*, a key TF regulating energy homeostasis (Duteil et al., 2010; Picard et al., 2002) and adipogenesis (Wang et al., 2008), showed co-association with a total of 326 genes, including relevant TF and genes associated with lipid and carbohydrate metabolisms, such as *PROX1*, *PBX1*, *ARNT*, *MYB*, *MTF2*, *TCF7L1*, *SCD5*, *ABCC2*, *INSIG1*, *ACACB*, *FABP4*, *FABP3*, *ME1*, *AASDH*, *ABCC5* and *SORT1*. A role for *PROX1* in the control of energy homeostasis has been proposed (Dufour et al., 2011). Moreover, association of SNPs mapped to *PROX1* and *SLC30A8* with fasting glucose levels and increased risk for T2D has been reported in humans (Dupuis et al., 2010). Both *PROX1* and *SLC30A8*, together with other T2D risk loci (*IL6R*, *TCF7L2*, *HNF1A*) and 21 genes reported as associated with plasma lipids in humans (Asselbergs et al., 2012) were predicted as target genes of *NCOA2* in our study. Co-expression analysis in adipose tissue validated 60.8% of the *NCOA2* co-association target genes, including *INSIG1* ($r_{\text{co-expression}} = 0.68$), *FDFT1* ($r_{\text{co-expression}} = 0.70$), *SETD2* ($r_{\text{co-expression}} = 0.59$) and *ABCC5* ($r_{\text{co-expression}} = 0.65$). In liver, 34.6% of the predicted

targets of *NCOA2* were validated, including the above-mentioned *PROX1* ($r_{\text{co-expression}} = 0.48$), *HNF1A* ($r_{\text{co-expression}} = 0.56$) and *TCF7L2* ($r_{\text{co-expression}} = 0.50$). It should be noted that previous studies in pigs show a correlation between *NCOA2* expression ($r = 0.605$, $P < 0.01$) and IMF content of LD muscle (Wang et al., 2008). Also, *NCOA2* was reported as modulating an AWM-network predicted for puberty in cattle (Fortes et al., 2011), which included fat deposition measurements as traits related to puberty. Furthermore, knockout *NCOA2*^{-/-} mice are protected against obesity, showing lean phenotype and decreased expression of genes involved in the uptake and storage of FA (Picard et al., 2002). A decreased expression of genes required for FA synthesis in liver tissue of *NCOA2*^{-/-} mice was observed (Jeong et al., 2006). In agreement with these previously results and the phenotypic difference in fat deposition between Iberian and Landrace breeds, a significant higher activity of *NCOA2* in the liver of Iberian pigs was detected (FC = 1.56, $P < 0.01$) relative to Landrace pigs (Figure 3). Along with the existing literature, our results support a key role of *NCOA2* in the regulation of FA in both humans and pigs alike.

Another TF predicted as critical for FA regulation was *EP300*, which encodes the adenovirus *E1A-associated cellular p300 transcriptional co-activator protein*. It functions as histone acetyltransferase that regulates transcription by chromatin remodelling. Via histone acetyltransferase activity, *EP300* regulates the transcription of liver X receptor (*LXR*) (Huuskonen et al., 2004). *EP300* is also required for adipocyte differentiation through the regulation of peroxisome proliferator-activated receptor gamma (*PPARG*) (Takahashi et al., 2002). Remarkably, *EP300* has been reported as transcriptional co-activator of estrogen receptor (*ER*), hepatocyte nuclear factor 4 α

(*HNF4-α*), aryl hydrocarbon receptor nuclear translocator (*ARNT*) and hepatocyte Nuclear Factor-1 α (*HNF1A*) (Ban et al., 2002; Chen et al., 2000; Torres-Padilla et al., 2003). All these above-mentioned TF co-regulated by *EP300* (*PPARG*, *LXR*, *HNF4*, *HNF1A*, *ER*, *ARNT*) influence lipid and carbohydrate metabolisms and have been extensively studied in this context (Alaynick, 2008; Huss et al., 2004; Li et al., 2004; Liu et al., 2012; Marcil et al., 2010; Palanker et al., 2009; Rankin et al., 2009a; Ulven et al., 2005; Xie et al., 2009). Among the 180 AWM-predicted target genes for *EP300*, there are 30 genes known to be involved in lipid metabolism including *ARNT* a member of the HIF-1 pathway. *ARNT* is a relevant TF regulating hepatic gluconeogenesis and lipogenic gene expression (Wang et al., 2009). Interestingly, we observed a significant co-expression between *ARNT* and *EP300* ($r = 0.61$) in the liver network. Additionally, other genes related to carbohydrate and lipid metabolism were predicted as *EP300* AWM-target genes. These included: *ADCY2*, *MMP9*, *ECHS1*, *ARRB1*, *EIF4E*, *ANK2*, *NR2E1*, *SLC2A6*, *SLC5A2*, *LEP*, *ELOVL6*, *MTTP*, *ACSM5*, *UCP2* and *CYP2E1* (for a full list see Supplementary Table ST7). Similarly to *NCOA2*, a significant higher expression of *EP300* in the liver of Iberian pigs was detected (FC = 1.23, $P < 0.05$) in comparison with Landrace pigs (Figure 3). Our results, predicting targets for *EP300* and studying their co-expression contributes to the knowledge on lipid and carbohydrate metabolism. It is well known that TF require co-regulators to modify and epigenetically remodel chromatin structure to facilitate the basal transcriptional machinery. *EP300* is a chromatin remodeling gene opening new possibilities to study the roll of epigenetic modifications in the regulation of pork meat quality and the molecular control of energy homeostasis.

The third key TF identified in our work was *FHL2*, an evolutionarily conserved gene that can interact with an important range of proteins from different functional classes, including receptors, signal transducers, TF and cofactors (Johannessen et al., 2006). *FHL2* plays an important role as molecular transmitter linking various signalling pathways to transcriptional regulation. For instance, *FHL2* is involved in the co-activation of human androgen receptor (*AR*), *ER* and peroxisome proliferator-activated receptor alpha (*PPAR α*) (Ciarlo et al., 2004; Johannessen et al., 2006; Muller et al., 2000). In addition, *FHL2* mediates interaction with β -catenin and promotes myoblast C2C12 differentiation in mice (Martin et al., 2002). The gene *B-cell CLL/lymphoma 9 (BCL9)*, an activator of the Wnt/ β -catenin (Brack et al., 2009) and *Wingless-type MMTV integration site family, member 4 (WNT4)* was among the 251 targets predicted for *FHL2* in our network. The growth factor *WNT4* is a member of the Wnt signaling pathway involved in developmental processes and relevant for gonad development and sex-determination (Bernard et al., 2007). Liver expression analyses provided supporting evidence for the predicted interaction between *FHL2* and *WNT4*, as a significant co-expression ($r = 0.44$, $P < 0.001$) was observed. Other genes and TF associated with development process, lipid and carbohydrate metabolism, such as *FHL5*, *MYO1E*, *MYB*, *RORC*, *JARID2*, *ZFHX4*, *WNK1*, *LIPC*, *CREB5*, *CDC42*, *ACSL1*, *FABP5*, *ABCB11*, *FLT1* and *HTR2A* were also predicted as targets of *FHL2* according to the co-association network. *FHL2* was not differentially expressed in the comparison between Iberian and Landrace pigs. Also, *FHL2* showed a proportion of validated interactions in the co-expression analysis (20% adipose tissues and 14.3% in liver) lower than for the other two TF, *NCOA2* and *EP300*. These somewhat less promising results could be a consequence

of the tissue-specific activity of *FHL2*, as it has been reported for the co-activation of *AR* (Muller et al., 2000). A regulatory impact of *FHL2* on FA pathways of pigs or humans could not be discarded, and needs further elucidation.

Although, some gene to gene interactions predicted by the AWM approach were not corroborated by the co-expression analysis, the possibility of these interactions occurring in other spatial temporal and/or tissues cannot be ruled out, or indeed manifesting their joint effect through other means than co-expression. TF and their target genes interact in a temporal and tissue dependent manner, so the examination of networks spanning multiple tissues is critical to highlight interactions that could otherwise be unknown from individual tissue analysis (Eisen et al., 1998). In spite of this tissue/time limitation, two of the three TF from the best trio (*EP300* and *NCOA2*) showed higher than average centrality values in both liver and adipose tissue co-expression networks. Moreover, we observed a significant co-expression between *NCOA2* and *EP300* in the liver network with some other TF considered master regulators of the lipid metabolism. For instance, *NCOA2* was significantly co-expressed with *PPAR α* ($r = 0.39$, $P < 0.01$), *HNF1A* ($r = 0.56$, $P < 0.001$) and *HNF4 α* ($r = 0.36$, $P < 0.01$), and *EP300* was co-expressed with *PPARD* ($r = 0.38$, $P < 0.01$) and *HNF1A* ($r = 0.64$, $P < 0.001$) (Supplementary Figure SF3 A, B). The liver plays a central role in maintaining overall energy balance by controlling lipid and carbohydrate metabolism. In pigs, the liver is the primary site of *de novo* cholesterol synthesis and fatty acid oxidation and, together with adipose tissue, has a crucial role in regulating lipid metabolism (Nafikov et al., 2007; O'Hea et al., 1969). All these observations, together with the higher expression of *NCOA2* and *EP300* observed in the liver of the Iberian pigs compared with Landrace

pigs, suggest a relevant role of these genes in the hepatic transcriptional regulation of lipid metabolism in pigs.

Overall, our GWAS and network predictions, supported by bibliography and co-expression analysis in liver and adipose tissue, suggest a co-operative role for the three TF (*NCOA2*, *EP300*, *FHL2*) in the transcriptional regulation of IMF, FA composition and the control of energy homeostasis in pigs. We hypothesize that these TF mediate a highly inter-connected regulatory cascade including pathways such as HIF-1, AR, ER and Wnt/ β -catenin that seem pivotal for lipid metabolism. In fact, the role of these pathways in the transcriptional regulation of lipid metabolism continues to be studied (Abiola et al., 2009; Alaynick, 2008; Goda et al., 2012; Huss et al., 2004; Rankin et al., 2009b; Willert et al., 2003; Zhou et al., 2012). A functional cooperation between the three TF in the modulation of these pathways is evident from our results and supported by literature evidence. For example, according to String database (Snel et al., 2000; Szklarczyk et al., 2011) (<http://string-db.org/>), experimental data confirmed that protein-protein interaction exists among, *EP300*, *NCOA2*, *FHL2*, *AR* and *ESR1* (Supplementary Figure SF4). In addition, *EP300* and *NCOA2* take part on the AR and ER pathways and both, *NCOA2* and *FHL2* are AR co-regulators (Heinlein et al., 2002; Muller et al., 2000; Urbanucci et al., 2008). Studying the combined effect of *NCOA2*, *EP300*, and *FHL2* in the regulation of specific genes will lead to new knowledge related to FA pathways.

The most overrepresented pathway corresponding to the 730 AWM-predicted target genes of the three TF was HIF-1 (Supplementary Figure SF2). The HIF-1 pathway is central to adaptive regulation of cellular energy metabolism; by regulating the expression of glycolytic enzymes and hepatic lipid metabolism (Goda et al., 2012;

Hamaguchi et al., 2008; Ke et al., 2006; Rankin et al., 2009b). Our liver co-expression analysis supports previously reported evidence (Wang et al., 2009) for the relevance of *ARNT* gene (member of HIF pathway) in the hepatic lipogenic gene expression. Additionally, HIF-1 α , which is another member of HIF pathway and β -catenin coordinately enhance AR transactivation. The interaction between β -catenin and both HIF-1 and AR pathways has been documented (Kaidi et al., 2007; Mitani et al., 2012; Yang et al., 2002). Moreover, β -catenin is a ligand-dependent co-activator of AR and a functional cooperation in the synergistic activation of AR-mediated transcription among *EP300*, *FHL2* and β -catenin have been reported (Labalette et al., 2004). Ours results showing the interactions between the three key TF, recapitulate these pathways interactions that are known mammalian biology, extending its significance to pigs.

In summary, our results suggest that common genetic variants mapped to (or in *linkage disequilibrium* with) *EP300*, *FHL2* and *NCOA2* together with other candidate genes including *ARNT*, *BCL9*, *SIRT1*, *PBX1*, *PROX1*, *HNFI1A*, *SLC30A8*, *TCF7L2* and *ANK2* modulate lipid metabolism and control energy homeostasis in pigs. Furthermore, epistatic predicted interactions between TF and their target genes are likely to contribute to the complex inheritance of FA composition and related polygenic traits (lipid metabolism and energy homeostasis). It is generally accepted that metabolic diseases such as obesity and T2D are linked to disturbance of energy homeostasis or homeostatic imbalance. It should be noted that among the 730 predicted target genes, an overrepresentation of genes from the T2D pathway was observed (Supplementary Figure SF2). Also, 39 of the 730 genes are known to control plasma lipid content in humans (Asselbergs et al., 2012; Teslovich et al., 2010).

Taken together, our results provide additional evidence supporting the adequacy of pig as an optimal animal model for the study of metabolic diseases in humans. Further studies will be required to elucidate the specific cellular and molecular processes of interaction among the three TF and its target genes that determine FA composition and control energy homeostasis in pigs. The implications of research in this area are broad, ranging from applications from pork meat quality to modeling human biology.

Methods

Phenotypic traits, animals and genotypes

Data from 144 pigs (25% Iberian × 75% Landrace), representing 26 full-sib families, from backcrossing five F1 males with 26 Landrace sows was utilized. Details about the management conditions and the phenotype information have been previously reported (Corominas et al., 2013; Ramayo-Caldas et al., 2012a; Ramayo-Caldas et al., 2012b). For this study and based on an previous principal components analysis (Ramayo-Caldas et al., 2012a) we selected 15 of the total 48 traits representing the most informative phenotypes within the dataset. Nine of the 15 traits were related to IMF fatty acid (FA) composition in LD muscle, seven correspond to indices of FA metabolism and the last one is the IMF percentage (Supplementary Table ST8). The Porcine SNP60K BeadChip (Illumina) (Ramos et al., 2009) was used to genotype a total 197 pigs, including the 144 phenotyped animals and the founder population. Quality control excluded SNPs with minor allele frequency < 5% and with call rate < 95%. A subset of 48,119 SNPs were retained for subsequent analysis, in addition, previously detected polymorphisms in the *MTTP* and *FABP4*, *FABP5* and *ELOVL6* genes were also tested (Corominas et al., 2013; Estelle et al., 2009; Mercade et al., 2006). The genomic

coordinates of the SNP correspond to the *Sus scrofa* genome sequence assembly (Sscrofa10.2, August 2011) (Groenen et al., 2012) and were annotated using as reference the pig assembly 10.2 [ftp://ftp.ncbi.nlm.nih.gov/genomes/Sus_scrofa/GFF/].

Statistical analysis

The GWAS was performed using Qxpack 5.0 software (Perez-Enciso et al., 2011). The additive effect of a SNP on each trait was estimate by mixed model (Henderson, 1984; Henderson, 1975) following the model:

$$y_{ij} = X\beta + Zu + s_{j,k} + e_{ij},$$

where: y_{ij} represents the vector of observations from the i^{th} pig at the j^{th} trait ; X is the incidence matrix relating fixed effects in β with observation in y_{ij} ; Z is the incidence matrix relating random additive polygenic effects in u with observation in y_{ij} ; $s_{j,k}$ represents the additive association of the k^{th} SNP on the j^{th} trait and e_{ij} is the vector of random residual effects. Fixed effects included in β were, sex (two levels), batch (five levels) and carcass weight as covariate. Polygenic effects were treated as random and distributed as $N(0, A\sigma_u)$ where A is a numerator of kinship matrix. Then, the allele substitution effect of the i^{th} SNP on the j^{th} trait was z-score standardized and employed to constructing the AWM (Fortes et al., 2010).

An R script was written to automate the process of building an AWM. Palmitoleic acid (C16:1 (n-7)) was use as the key phenotype and we followed the procedure describe by Fortes and colleagues (Fortes et al., 2010), but introduced a few modifications, specifically regarding the P -value threshold for selecting SNP from GWAS. The P -value threshold was chosen by exploring the sensitivity of the data instead of simply accepting

the nominal $P < 0.05$. We took advantage of the biological knowledge concerning TF related to the analyzed traits and used it as *a priori* information. The process for choosing the threshold was as follows:

- **Step1:** A total of 340 TF were located within 2.5Kb of a SNP and therefore included in the initial dataset. For all these TF included in our dataset, those that are well known key regulators of the lipid metabolism were initially selected.
- **Step2:** For each gene, those involved in the lipid metabolism and also reported in the census of human TF by Vaqueriza et al (Vaquerizas et al., 2009) were included.
- **Step3:** The Human Protein Reference Database (HPRD) and the Biomolecular Object Network Databank (BIND) were mined. Then other TF that have been reported to interact with some of the TF retained in the two previous steps were selected. After these first 3 steps, a total of 34 TF were retained (Supplementary table ST9).
- **Step4:** Subsequently we compare the distribution of the 34 TF at different P -values from $P=1$ to $P=10^{-4}$ versus the distribution of the total number of TF included in the AWM (340). As a result, we have chosen $P < 0.035$ as the threshold. This specific P -value maximizes the difference between both groups of TF (Figure 9), imposing an informed bias towards lipid metabolism to the network.

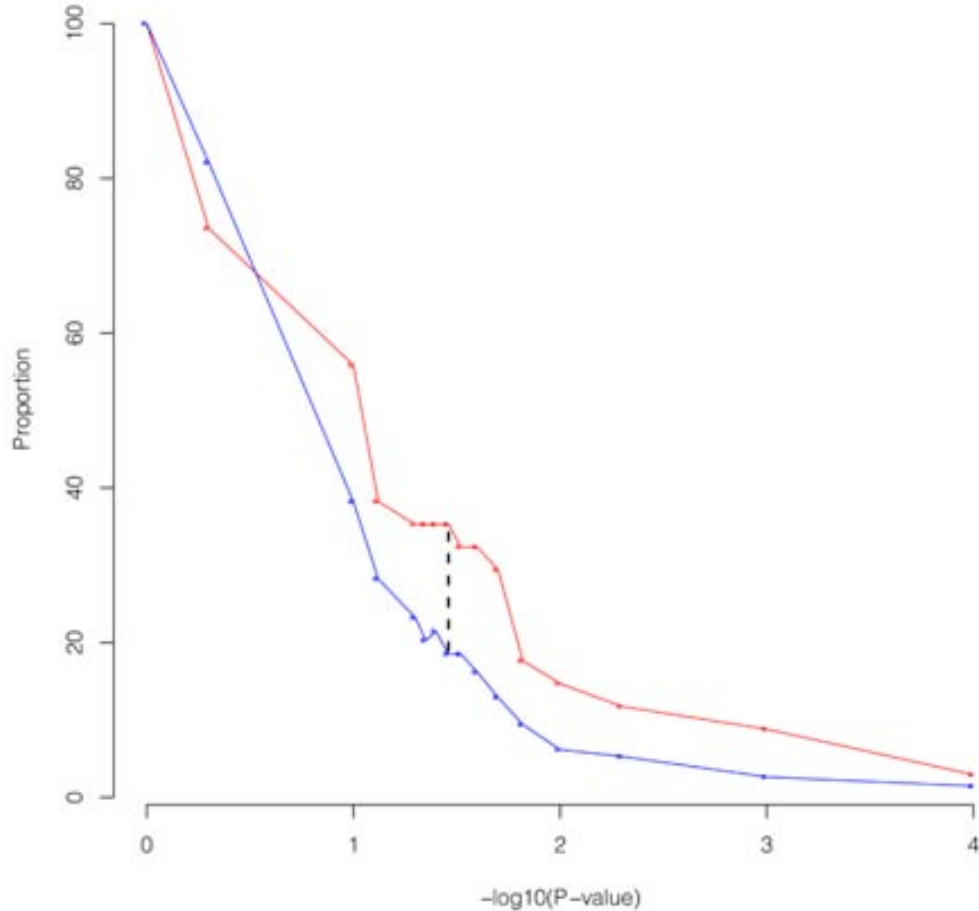


Figure 9. Sensitivity analysis of the 34 lipid-related TF at different P -values (from $P < 1$ to $P < 10^{-4}$) against the distribution of the total number of TF included in the dataset.

After defining the threshold of $P < 0.035$, the selection of SNPs for building AWM continued. Those SNPs that were either associated ($P < 0.035$) with palmitoleic acid or with any ≥ 3 traits, and were located either $\leq 2,500$ bp to or ≥ 850 kb from the nearest annotated gene (Sscrofa10.2 assembly), were selected to build the AWM matrix. PermutMatrix software (Caraux et al., 2005) was employed to visualize hierarchical clustering of traits (AWM columns) and genes (AWM rows). To identify and report

gene-gene or gene-SNP interactions we used PCIT algorithm (Reverter et al., 2008). Cytoscape software (Shannon et al., 2003) was used to visualize the gene network and also to perform overrepresented GO terms analysis, using BiNGO plugin (Maere et al., 2005). Pathway enrichment analysis were performed using FATIGO tool from BABELOMICS (Al-Shahrour et al., 2007; Medina et al., 2010). Further, pathways analyses of the 730 predicted target genes (co-associated with the key TF) were performed using ClueGO, Cytoscape plugin (Bindea et al., 2009). Pathway information was retrieved from the KEGG (<http://www.genome.jp/kegg/>) and BioCarta (<http://www.biocarta.com/>) databases. In all cases, the cut-off for considering a significance overrepresentation was established by multiple testing correction of the *P*-value (FDR < 0.05).

Expression and co-expression analysis

In order to provide supporting evidence for the *in-silico* AWM-network predictions we obtained and explored gene expression data by reverse transcription quantitative Real-Time PCR (RT-qPCR). First, the expression pattern of the 3 Key TF (*NCOA2*, *EP300*, *FHL2*) in LD muscle, liver and adipose tissues was tested in two phenotypically divergent breeds for fat deposition traits (Iberian and Landrace which are also the founders of our studied population, five animals per breed). Second, we compared the gene expression patterns between Iberian and Landrace pigs of the 3 TF across all tissues. Finally, liver and adipose co-expression analyses of 55 (43 from the present study, and twelve: *ACSM5*, *APOA2*, *ARNT*, *CYP7A1*, *FABP5*, *FADS3*, *HNF4a*, *LIPC*, *MTTP*, *PPARA*, *PPARD* and *ELOVL6* genes from Ballester et al., 2013 Submitted) and 40 genes, respectively, were performed using the PCIT algorithm (Reverter et al.,

2008) in 55 backcross animals. Since sex differences in liver transcriptome have been reported (Zhang et al., 2011) only females were considered in the co-expression analyses of both tissues.

From the 55 genes explored in the liver co-expression analysis, 48 were present in the AWM network. The remaining seven were incorporated due to their biological relevance, including three well-known TF related to lipid metabolism (*PPAR α* , *PPARD*, *HNF4 α*) and four genes related to lipid metabolism (*SIRT1*, *FADS3*, *APOA2*, *CYP7A1*). Similarly, from the 40 genes employed in the adipose co-expression analysis, 39 were present in the AWM network. The one gene out, *SIRT*, was also included due to its relevant controlling lipolysis (Chakrabarti et al., 2011; Schug et al., 2011) and promoting fat mobilization in white adipose tissue (Picard et al., 2004).

Total RNA was obtained from liver, muscle and adipose tissues using the RiboPure kit (Ambion), following the manufacturer's recommendations. RNA was quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop products) and the RNA integrity was assessed by Agilent Bioanalyzer-2100 (Agilent Technologies). Approximately, one microgram of total RNA was reverse-transcribed into cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) in 20 μ l of reactions, following the manufacturer's instructions.

To analyze the expression pattern of the 3 key transcription factors, an ABI PRISM 7900 Sequence Detection System (Applied Biosystems) in combination with FastStart Universal Sybr green master (Rox; Roche Applied Science) was used. PCR amplifications were performed in a total reaction volume of 20 μ l containing 5 μ l of cDNA diluted 1:25. All primers were used at 300 nM. The thermal cycle was 10 min at

95°C, 40 cycles of 15s at 95°C and 1 min at 60°C. A dissociation curve was drawn for each primer pair to assess the specificity of the amplification. Three reference genes (*ACTB*, *HPRT1*, *TBP*) frequently used in RT-qPCR experiments were tested as endogenous controls. Using the GeNorm software (Vandesompele et al., 2002), the *ACTB* and *TBP* genes were selected as the best endogenous controls for all tissues. After ensuring the possibility to use the $2^{-\Delta\Delta CT}$ method (Livak et al., 2001), data was analyzed using the RQ manager v1.2.1 and the DataAssist™v3.0 softwares (Applied Biosystems). The $2^{-\Delta CT}$ values were used to compare our data.

The 48.48 microfluidic dynamic array IFC chip (Fluidigm) was used to analyze the expression of 48 genes (44 target genes and 4 reference genes) in liver and adipose tissue of 55 backcross animals. Two μ l of 1:5 diluted cDNA was pre-amplified using 2X Taqman PreAmp Master Mix (Applied Biosystems) and 50nM of each primer pair in 5 μ l reaction volume, according to the manufacturer's directions. The cycling program was 10 min at 95°C followed by 16 cycles of 15s at 95°C and 4 min at 60°C. At the end of this pre-amplification step, the reactions were diluted 1:5 (diluted pre-amplification samples). RT-qPCR on the dynamic array chips was conducted on the BioMark™ system (Fluidigm). Five μ l sample pre-mix containing 2.5 μ l of SsoFast EvaGreen Supermix with Low ROX (Bio-Rad), 0.25 μ l of DNA Binding Dye Sample Loading Reagent (Fluidigm) and 2.25 μ l of diluted pre-amplification samples (1:16 or 1:64 from the diluted pre-amplification samples from liver and backfat, respectively), as well as 5 μ l assay mix containing 2.5 μ l of Assay Loading Reagent (Fluidigm), 2.25 μ l of DNA Suspension Buffer (Teknova) and 0.25 μ l of 100 μ M primer pairs (500nM in the final reaction) were mixed inside the chip using the IFC controller MX (Fluidigm). The thermal cycle was

60s at 95°C followed by 30 cycles of 5s at 96°C and 20s at 60°C. A dissociation curve was also drawn for each primer pair.

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, submitted) applying the relative standard curve method (see Applied Biosystems user bulletin #2). Standard curves with a four-fold dilutions series (1/4, 1/16, 1/64, 1/256, 1/1024) of a pool of 10 cDNA samples were constructed for each gene to extrapolate the quantity values of the studied samples. The PCR efficiencies were almost 100% in both tissues for all the assays (Supplementary table ST10) with low coefficients of inter-assay variation of threshold cycle (<2.4% in liver and <3.5% in adipose tissue). Of the four endogenous genes tested (*ACTB*, *B2M*, *HPRT1*, *TBP*), *ACTB* and *TBP* were the genes with the most stable expression (Vandesompele et al., 2002) in both tissues. The normalized quantity values of each sample and assay were used to compare our data.

All the primers used in this study were designed using PrimerExpress 2.0 software (Applied Biosystems) and are shown in Supplementary table ST10. Prior to perform the Fluidigm Real-Time PCR, all the assays were tested for PCR specificity in an ABI PRISM 7900 Sequence Detection System (Applied Biosystems) using two-fold dilutions (1/20, 1/200) of a pool of ten cDNA samples and a minus RT control to check the presence of DNA contamination. Melting curve analysis was performed for all the assays.

Acknowledgments

This work was funded by MICINN project AGL2011-29821-C02 (Ministerio de Economía y Competitividad), and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036, Centre for Research in Agrigenomics), we are grateful to Dr. Nick Hudson for insightful suggestions and manuscript proofread. Y. Ramayo-Caldas was funded by a FPU PhD grant from the Spanish Ministerio de Educación (AP2008-01450).

References

- Abiola, M., M. Favier, E. Christodoulou-Vafeiadou, A.-L. Pichard, I. Martelly, and I. Guillet-Deniau. 2009. Activation of wnt/ β -catenin signaling increases insulin sensitivity through a reciprocal regulation of wnt10b and srebp-1c in skeletal muscle cells. *PLoS ONE* 4: e8509.
- Al-Shahrour, F., P. Minguéz, J. Tarraga, I. Medina, E. Alloza, D. Montaner, and J. Dopazo. 2007. Fatigo +: A functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Research* 35: W91-W96.
- Alaynick, W. A. 2008. Nuclear receptors, mitochondria and lipid metabolism. *Mitochondrion* 8: 329-337.
- Asselbergs, F. W., Y. Guo, E. P. A. van Iperen, S. Sivapalaratnam, V. Tragante, M. B. Lanktree, L. A. Lange, B. Almoguera, Y. E. Appelman, J. Barnard, J. Baumert, A. L. Beitelshes, T. R. Bhangale, Y.-D. I. Chen, T. R. Gaunt, Y. Gong, J. C. Hopewell, T. Johnson, M. E. Kleber, T. Y. Langaee, M. Li, Y. R. Li, K. Liu, C. W. McDonough, M. F. L. Meijs, R. P. S. Middelberg, K. Musunuru, C. P. Nelson, J. R. O'Connell, S. Padmanabhan, J. S. Pankow, N. Pankratz, S. Rafelt, R. Rajagopalan, S. P. R. Romaine, N. J. Schork, J. Shaffer, H. Shen, E. N. Smith, S. E. Tischfield, P. J. van der Most, J. V. van Vliet-Ostaptchouk, N. Verweij, K. A. Volcik, L. Zhang, K. R. Bailey, K. M. Bailey, F. Bauer, J. M. A. Boer, P. S. Braund, A. Burt, P. R. Burton, S. G. Buxbaum, W. Chen, R. M. Cooper-DeHoff,

L. A. Cupples, J. S. deJong, C. Delles, D. Duggan, M. Fornage, C. E. Furlong, N. Glazer, J. G. Gums, C. Hastie, M. V. Holmes, T. Illig, S. A. Kirkland, M. Kivimaki, R. Klein, B. E. Klein, C. Kooperberg, K. Kottke-Marchant, M. Kumari, A. Z. LaCroix, L. Mallela, G. Murugesan, J. Ordovas, W. H. Ouwehand, W. S. Post, R. Saxena, H. Scharnagl, P. J. Schreiner, T. Shah, D. C. Shields, D. Shimbo, S. R. Srinivasan, R. P. Stolk, D. I. Swerdlow, H. A. Taylor Jr, E. J. Topol, E. Toskala, J. L. van Pelt, J. van Setten, S. Yusuf, J. C. Whittaker, A. H. Zwinderman, S. S. Anand, A. J. Balmforth, G. S. Berenson, C. R. Bezzina, B. O. Boehm, E. Boerwinkle, J. P. Casas, M. J. Caulfield, R. Clarke, J. M. Connell, K. J. Cruickshanks, K. W. Davidson, I. N. M. Day, P. I. W. de Bakker, P. A. Doevendans, A. F. Dominiczak, A. S. Hall, C. A. Hartman, C. Hengstenberg, H. L. Hillege, M. H. Hofker, S. E. Humphries, G. P. Jarvik, J. A. Johnson, B. M. Kaess, S. Kathiresan, W. Koenig, D. A. Lawlor, W. M. M. Verschuren, O. Melander, B. D. Mitchell, G. W. Montgomery, P. B. Munroe, S. S. Murray, S. J. Newhouse, N. C. Onland-Moret, N. Poulter, B. Psaty, S. Redline, S. S. Rich, J. I. Rotter, H. Schunkert, P. Sever, A. R. Shuldiner, R. L. Silverstein, A. Stanton, B. Thorand, M. D. Trip, M. Y. Tsai, P. van der Harst, E. van der Schoot, Y. T. van der Schouw, W. M. M. Verschuren, H. Watkins, A. A. M. Wilde, B. H. R. Wolffenbuttel, J. B. Whitfield, G. K. Hovingh, C. M. Ballantyne, C. Wijmenga, M. P. Reilly, N. G. Martin, J. G. Wilson, D. J. Rader, N. J. Samani, A. P. Reiner, R. A. Hegele, J. J. P. Kastelein, A. D. Hingorani, P. J. Talmud, H. Hakonarson, C. C. Elbers, B. J. Keating, and F. Drenos. 2012. Large-scale gene-

- centric meta-analysis across 32 studies identifies multiple lipid loci. *The American Journal of Human Genetics* 91: 823-838.
- Ballester, M., R. Cordón, and J. M. Folch. 2013. Dag expression: High-throughput gene expression analysis of real time pcr data using standard curves for relative quantification. Submitted.
- Ban, N., Y. Yamada, Y. Someya, K. Miyawaki, Y. Ihara, M. Hosokawa, S. Toyokuni, K. Tsuda, and Y. Seino. 2002. Hepatocyte nuclear factor-1 α recruits the transcriptional co-activator p300 on the glut2 gene promoter. *Diabetes* 51: 1409-1418.
- Bendixen, E., M. Danielsen, K. Larsen, and C. Bendixen. 2010. Advances in porcine genomics and proteomics--a toolbox for developing the pig as a model organism for molecular biomedical research. *Brief Funct Genomics* 9: 208-219.
- Bernard, P., and V. R. Harley. 2007. Wnt4 action in gonadal development and sex determination. *The International Journal of Biochemistry & Cell Biology* 39: 31-43.
- Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.-H. Fridman, F. Pagés, Z. Trajanoski, and J. r. m. Galon. 2009. Cluego: A cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25: 1091-1093.
- Brack, A. S., F. Murphy-Seiler, J. Hanifi, J. r. Deka, S. Eyckerman, C. Keller, M. Aguet, and T. A. Rando. 2009. Bcl9 is an essential component of canonical wnt signaling that mediates the differentiation of myogenic progenitors during muscle regeneration. *Developmental Biology* 335: 93-105.

- Caraux, G., and S. Pinloche. 2005. Permutmatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 21: 1280-1281.
- Casellas, J., J. L. Noguera, J. Reixach, I. Diaz, M. Amills, and R. Quintanilla. 2010. Bayes factor analyses of heritability for serum and muscle lipid traits in duroc pigs. *J Anim Sci* 88: 2246-2254.
- Chakrabarti, P., T. English, S. Karki, L. Qiang, R. Tao, J. Kim, Z. Luo, S. R. Farmer, and K. V. Kandror. 2011. Sirt1 controls lipolysis in adipocytes via foxo1-mediated expression of atgl. *Journal of Lipid Research* 52: 1693-1701.
- Chen, D., S.-M. Huang, and M. R. Stallcup. 2000. Synergistic, p160 coactivator-dependent enhancement of estrogen receptor function by carm1 and p300. *Journal of Biological Chemistry* 275: 40810-40816.
- Ciarlo, J. D., A. M. Flores, N. G. McHugh, and B. J. Aneskievich. 2004. Fhl2 expression in keratinocytes and transcriptional effect on ppar α /rxr α . *Journal of dermatological science* 35: 61-63.
- Clop, A., C. Ovilo, M. Perez-Enciso, A. Cercos, A. Tomas, A. Fernandez, A. Coll, J. M. Folch, C. Barragan, I. Diaz, M. A. Oliver, L. Varona, L. Silio, A. Sanchez, and J. L. Noguera. 2003. Detection of qtl affecting fatty acid composition in the pig. *Mamm Genome* 14: 650-656.
- Corominas, J., Y. Ramayo-Caldas, A. Puig-Oliveras, D. Perez-Montarelo, J. L. Noguera, J. M. Folch, and M. Ballester. 2013. Polymorphism in the elovl6 gene is associated with a major qtl effect on fatty acid composition in pigs. *PLoS ONE* 8: e53687.

Dufour, C. R., M.-P. Levasseur, N. H. H. Pham, L. J. Eichner, B. J. Wilson, A. Charest-Marcotte, D. Duguay, J.-F. o. Poirier-H[√]on, N. Cermakian, and V. Gigu[√]re. 2011. Genomic convergence among *errα*, *prox1*, and *bmal1* in the control of metabolic clock outputs. *PLoS Genet* 7: e1002143.

Dupuis, J., C. Langenberg, I. Prokopenko, R. Saxena, N. Soranzo, A. U. Jackson, E. Wheeler, N. L. Glazer, N. Bouatia-Naji, A. L. Gloyn, C. M. Lindgren, R. Magi, A. P. Morris, J. Randall, T. Johnson, P. Elliott, D. Rybin, G. Thorleifsson, V. Steinthorsdottir, P. Henneman, H. Grallert, A. Dehghan, J. J. Hottenga, C. S. Franklin, P. Navarro, K. Song, A. Goel, J. R. B. Perry, J. M. Egan, T. Lajunen, N. Grarup, T. Sparso, A. Doney, B. F. Voight, H. M. Stringham, M. Li, S. Kanoni, P. Shrader, C. Cavalcanti-Proenca, M. Kumari, L. Qi, N. J. Timpson, C. Gieger, C. Zabena, G. Rocheleau, E. Ingelsson, P. An, J. O'Connell, J. a. Luan, A. Elliott, S. A. McCarroll, F. Payne, R. M. Roccascocca, F. Pattou, P. Sethupathy, K. Ardlie, Y. Ariyurek, B. Balkau, P. Barter, J. P. Beilby, Y. Ben-Shlomo, R. Benediktsson, A. J. Bennett, S. Bergmann, M. Bochud, E. Boerwinkle, A. Bonnefond, L. L. Bonnycastle, K. Borch-Johnsen, Y. Bottcher, E. Brunner, S. J. Bumpstead, G. Charpentier, Y.-D. I. Chen, P. Chines, R. Clarke, L. J. M. Coin, M. N. Cooper, M. Cornelis, G. Crawford, L. Crisponi, I. N. M. Day, E. J. C. de Geus, J. Delplanque, C. Dina, M. R. Erdos, A. C. Fedson, A. Fischer-Rosinsky, N. G. Forouhi, C. S. Fox, R. Frants, M. G. Franzosi, P. Galan, M. O. Goodarzi, J. Graessler, C. J. Groves, S. Grundy, R. Gwilliam, U. Gyllensten, S. Hadjadj, G. Hallmans, N. Hammond, X. Han, A.-L. Hartikainen, N. Hassanali, C. Hayward, S. C. Heath, S. Hercberg, C. Herder, A. A. Hicks, D. R. Hillman, A. D. Hingorani, A. Hofman, J.

Hui, J. Hung, B. Isomaa, P. R. V. Johnson, T. Jorgensen, A. Jula, M. Kaakinen, J. Kaprio, Y. A. Kesaniemi, M. Kivimaki, B. Knight, S. Koskinen, P. Kovacs, K. O. Kyvik, G. M. Lathrop, D. A. Lawlor, O. Le Bacquer, C. Lecoeur, Y. Li, V. Lyssenko, R. Mahley, M. Mangino, A. K. Manning, M. T. Martinez-Larrad, J. B. McAteer, L. J. McCulloch, R. McPherson, C. Meisinger, D. Melzer, D. Meyre, B. D. Mitchell, M. A. Morken, S. Mukherjee, S. Naitza, N. Narisu, M. J. Neville, B. A. Oostra, M. Orru, R. Pakyz, C. N. A. Palmer, G. Paolisso, C. Pattaro, D. Pearson, J. F. Peden, N. L. Pedersen, M. Perola, A. F. H. Pfeiffer, I. Pichler, O. Polasek, D. Posthuma, S. C. Potter, A. Pouta, M. A. Province, B. M. Psaty, W. Rathmann, N. W. Rayner, K. Rice, S. Ripatti, F. Rivadeneira, M. Roden, O. Rolandsson, A. Sandbaek, M. Sandhu, S. Sanna, A. A. Sayer, P. Scheet, L. J. Scott, U. Seedorf, S. J. Sharp, B. Shields, G. Sigursson, E. J. G. Sijbrands, A. Silveira, L. Simpson, A. Singleton, N. L. Smith, U. Sovio, A. Swift, H. Syddall, A.-C. Syvanen, T. Tanaka, B. Thorand, J. Tichet, A. Tonjes, T. Tuomi, A. G. Uitterlinden, K. W. van Dijk, M. van Hoek, D. Varma, S. Visvikis-Siest, V. Vitart, N. Vogelzangs, G. Waeber, P. J. Wagner, A. Walley, G. B. Walters, K. L. Ward, H. Watkins, M. N. Weedon, S. H. Wild, G. Willemsen, J. C. M. Witteman, J. W. G. Yarnell, E. Zeggini, D. Zelenika, B. Zethelius, G. Zhai, J. H. Zhao, M. C. Zillikens, I. B. Borecki, R. J. F. Loos, P. Meneton, P. K. E. Magnusson, D. M. Nathan, G. H. Williams, A. T. Hattersley, K. Silander, V. Salomaa, G. D. Smith, S. R. Bornstein, P. Schwarz, J. Spranger, F. Karpe, A. R. Shuldiner, C. Cooper, G. V. Dedoussis, M. Serrano-Rios, A. D. Morris, L. Lind, L. J. Palmer, F. B. Hu, P. W. Franks, S. Ebrahim, M. Marmot, W. H. L. Kao, J. S. Pankow, M. J. Sampson,

- J. Kuusisto, M. Laakso, T. Hansen, O. Pedersen, P. P. Pramstaller, H. E. Wichmann, T. Illig, I. Rudan, A. F. Wright, M. Stumvoll, H. Campbell, and J. F. Wilson. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42: 105-116.
- Duteil, D., C. Chambon, F. Ali, R. Malivindi, J. Zoll, S. Kato, B. Geny, P. Chambon, and D. Metzger. 2010. The transcriptional coregulators *tif2* and *src-1* regulate energy homeostasis by modulating mitochondrial respiration in skeletal muscles. *Cell Metab* 12: 496-508.
- Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863-14868.
- Ekser, B., M. Ezzelarab, H. Hara, D. J. van der Windt, M. Wijkstrom, R. Bottino, M. Trucco, and D. K. Cooper. 2012. Clinical xenotransplantation: The next medical revolution? *Lancet* 379: 672-683.
- Estelle, J., A. I. Fernandez, M. Perez-Enciso, A. Fernandez, C. Rodriguez, A. Sanchez, J. L. Noguera, and J. M. Folch. 2009. A non-synonymous mutation in a conserved site of the *mtp* gene is strongly associated with protein activity and fatty acid profile in pigs. *Anim Genet* 40: 813-820.
- Esteve, A., A. Ojeda, L. S. Huang, J. M. Folch, and M. Perez-Enciso. 2011. Nucleotide variability of the porcine *serpina6* gene and the origin of a putative causal mutation associated with meat quality. *Anim Genet* 42: 235-241.
- Fernandez, A. I., D. Perez-Montarelo, C. Barragan, Y. Ramayo-Caldas, N. Ibanez-Escriche, A. Castello, J. L. Noguera, L. Silio, J. M. Folch, and M. C. Rodriguez.

2012. Genome-wide linkage analysis of qtl for growth and body composition employing the porcinesnp60 beadchip. *BMC Genet* 13: 41.
- Fortes, M. R., A. Reverter, Y. Zhang, E. Collis, S. H. Nagaraj, N. N. Jonsson, K. C. Prayaga, W. Barris, and R. J. Hawken. 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. *Proc Natl Acad Sci U S A* 107: 13642-13647.
- Fortes, M. R. S., A. Reverter, S. H. Nagaraj, Y. Zhang, N. N. Jonsson, W. Barris, S. Lehnert, G. B. Boe-Hansen, and R. J. Hawken. 2011. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. *Journal of Animal Science* 89: 1669-1683.
- Goda, N., and M. Kanai. 2012. Hypoxia-inducible factors and their roles in energy metabolism. *International Journal of Hematology* 95: 457-463.
- Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H.-J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. F. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S.-H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. M. A. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapia, B. Dibbits, N. Drou, Z.-Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. R. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z.-L. Hu, S. J. Humphray, T. Hunt, H. Hornshoj, J.-T. Jeon, P. Jern, M. Jones, J. Jurka,

- H. Kanamori, R. Kapetanovic, J. Kim, J.-H. Kim, K.-W. Kim, T.-H. Kim, G. Larson, K. Lee, K.-T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. Truong Nguyen, P. Ni, S.-J. Oh, S. Onteru, F. Panitz, E.-W. Park, H.-S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, and L. B. Schook. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393-398.
- Hamaguchi, T., N. Iizuka, R. Tsunedomi, Y. Hamamoto, T. Miyamoto, M. Iida, Y. Tokuhisa, K. Sakamoto, M. Takashima, T. Tamesa, and M. Oka. 2008. Glycolysis module activated by hypoxia-inducible factor 1alpha is related to the aggressive phenotype of hepatocellular carcinoma. *Int J Oncol* 33: 725-731.
- Hardie, D. G. 2012. Organismal carbohydrate and lipid homeostasis. *Cold Spring Harb Perspect Biol* 4.
- Heinlein, C., and C. Chang. 2002. Androgen receptor (ar) coregulators: An overview. *Endocr Rev* 23: 175 - 200.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31: 423-447.

- Henderson, C. R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario, Canada.
- Hernandez-Sanchez, J., M. Amills, R. N. Pena, A. Mercade, A. Manunza, and R. Quintanilla. 2012. Genomic architecture of heritability and genetic correlations for intramuscular and back fat contents in duroc pigs. *Journal of Animal Science*.
- Haupt, K., T. Haupt, and W. Pond. 1979. The pig as a model for the study of obesity and of control of food intake: A review. *Yale J Biol Med* May-Jun; 52(3): 307–329.
- Huss, J. M., I. P. Torra, B. Staels, V. Giguere, and D. P. Kelly. 2004. Estrogen-related receptor α directs peroxisome proliferator-activated receptor α signaling in the transcriptional control of energy metabolism in cardiac and skeletal muscle. *Molecular and Cellular Biology* 24: 9079-9091.
- Huuskonen, J., P. E. Fielding, and C. J. Fielding. 2004. Role of p160 coactivator complex in the activation of liver x receptor. *Arteriosclerosis, Thrombosis, and Vascular Biology* 24: 703-708.
- Jeong, J.-W., I. Kwak, K. Y. Lee, L. D. White, X.-P. Wang, F. C. Brunicardi, B. W. O, Malley, and F. J. DeMayo. 2006. The genomic analysis of the impact of steroid receptor coactivators ablation on hepatic metabolism. *Molecular Endocrinology* 20: 1138-1152.
- Johannessen, M., S. Moller, T. Hansen, U. Moens, and M. V. Ghelue. 2006. The multifunctional roles of the four-and-a-half-lim only protein fhl2. *Cellular and Molecular Life Sciences CMLS* 63: 268-284.
- Kaidi, A., A. C. Williams, and C. Paraskeva. 2007. Interaction between β -catenin and hif-1 promotes cellular adaptation to hypoxia. *Nat Cell Biol* 9: 210-217.

- Ke, Q., and M. Costa. 2006. Hypoxia-inducible factor-1 (hif-1). *Mol Pharmacol* 70: 1469-1480.
- Labalette, C., C.-A. I. Renard, C. Neuveut, M.-A. Buendia, and Y. Wei. 2004. Interaction and functional cooperation between the lim protein fhl2, cbp/p300, and β -catenin. *Molecular and Cellular Biology* 24: 10689-10702.
- Larson, G., K. Dobney, U. Albarella, M. Fang, E. Matisoo-Smith, J. Robins, S. Lowden, H. Finlayson, T. Brand, E. Willerslev, P. Rowley-Conwy, L. Andersson, and A. Cooper. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307: 1618-1621.
- Li, A. C., and C. K. Glass. 2004. Ppar- and lxr-dependent pathways controlling lipid metabolism and the development of atherosclerosis. *Journal of Lipid Research* 45: 2161-2173.
- Liu, Y., D. K. Qiu, and X. Ma. 2012. Liver x receptors bridge hepatic lipid metabolism and inflammation. *Journal of Digestive Diseases* 13: 69-74.
- Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative pcr and the 2- $\delta\delta$ ct method. *Methods* 25: 402-408.
- Maere, S., K. Heymans, and M. Kuiper. 2005. Bingo: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448-3449.
- Marcil, V., E. Seidman, D. Sinnett, F. Boudreau, F.-P. Gendron, J.-F. Beaulieu, D. Menard, L.-P. Precourt, D. Amre, and E. Levy. 2010. Modification in oxidative stress, inflammation, and lipoprotein assembly in response to hepatocyte nuclear

- factor 4 α ± knockdown in intestinal epithelial cells. *Journal of Biological Chemistry* 285: 40448-40460.
- Martin, B., R. Schneider, S. Janetzky, Z. Waibler, P. Pandur, M. Kuhl, J. Behrens, K. von der Mark, A. Starzinski-Powitz, and V. Wixler. 2002. The lim-only protein fhl2 interacts with β -catenin and promotes differentiation of mouse myoblasts. *The Journal of Cell Biology* 159: 113-122.
- Medina, I., J. Carbonell, L. Pulido, S. C. Madeira, S. Goetz, A. Conesa, J. n. Tárraga, A. Pascual-Montano, R. Nogales-Cadenas, J. Santoyo, F. García, M. Marbá, D. Montaner, and J. n. Dopazo. 2010. Babelomics: An integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Research* 38: W210-W213.
- Mercade, A., M. Perez-Enciso, L. Varona, E. Alves, J. L. Noguera, A. Sanchez, and J. M. Folch. 2006. Adipocyte fatty-acid binding protein is closely associated to the porcine fat1 locus on chromosome 4. *Journal of Animal Science* 84: 2907-2913.
- Mitani, T., N. Harada, Y. Nakano, H. Inui, and R. Yamaji. 2012. Coordinated action of hypoxia-inducible factor-1 α and β -catenin in androgen receptor signaling. *Journal of Biological Chemistry* 287: 33594-33606.
- Mote, B. E., K. J. Koehler, J. W. Mabry, K. J. Stalder, and M. F. Rothschild. 2009. Identification of genetic markers for productive life in commercial sows. *J Anim Sci* 87: 2187-2195.
- Muller, J. M., U. Isele, E. Metzger, A. Rempel, M. Moser, A. Pscherer, T. Breyer, C. Holubarsch, R. Buettner, and R. Schule. 2000. Fhl2, a novel tissue-specific coactivator of the androgen receptor. *EMBO J* 19: 359-369.

- Nafikov, R. A., and D. C. Beitz. 2007. Carbohydrate and lipid metabolism in farm animals. *The Journal of Nutrition* 137: 702-705.
- Ntawubizi, M., E. Colman, S. Janssens, K. Raes, N. Buys, and S. De Smet. 2010. Genetic parameters for intramuscular fatty acid composition and metabolism in pigs. *J Anim Sci* 88: 1286-1294.
- O'Hea, E. K., and G. A. Leveille. 1969. Significance of adipose tissue and liver as sites of fatty acid synthesis in the pig and the efficiency of utilization of various substrates for lipogenesis. *The Journal of Nutrition* 99: 338-344.
- Palanker, L., J. M. Tennessen, G. Lam, and C. S. Thummel. 2009. *Drosophila hnf4* regulates lipid mobilization and β -oxidation. *Cell Metabolism* 9: 228-239.
- Parks, E. J., R. M. Krauss, M. P. Christiansen, R. A. Neese, and M. K. Hellerstein. 1999. Effects of a low-fat, high-carbohydrate diet on vldl-triglyceride assembly, production, and clearance. *J Clin Invest* 104(8): 1087-1096.
- Perez-Enciso, M., A. Clop, J. L. Noguera, C. Ovilo, A. Coll, J. M. Folch, D. Babot, J. Estany, M. A. Oliver, I. Diaz, and A. Sanchez. 2000. A qtl on pig chromosome 4 affects fatty acid metabolism: Evidence from an iberian by landrace intercross. *J Anim Sci* 78: 2525-2531.
- Perez-Enciso, M., and I. Misztal. 2011. Qxpk.5: Old mixed model solutions for new genomics problems. *BMC Bioinformatics* 12: 202.
- Picard, F., M. Gehin, J.-S. Annicotte, S. Rocchi, M.-F. Champy, B. W. O'Malley, P. Chambon, and J. Auwerx. 2002. Src-1 and tif2 control energy balance between white and brown adipose tissues. *Cell* 111: 931-941.

- Picard, F., M. Kurtev, N. Chung, A. Topark-Ngarm, T. Senawong, R. Machado De Oliveira, M. Leid, M. W. McBurney, and L. Guarente. 2004. Sirt1 promotes fat mobilization in white adipocytes by repressing ppar-gamma. *Nature* 429: 771-776.
- Ramayo-Caldas, Y., N. Mach, A. Esteve-Codina, J. Corominas, A. Castello, M. Ballester, J. Estelle, N. Ibanez-Escriche, A. I. Fernandez, M. Perez-Enciso, and J. M. Folch. 2012a. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics* 13: 547.
- Ramayo-Caldas, Y., A. Mercade, A. Castello, B. Yang, C. Rodriguez, E. Alves, I. Diaz, N. Ibanez-Escriche, J. L. Noguera, M. Perez-Enciso, A. I. Fernandez, and J. M. Folch. 2012b. Genome-wide association study for intramuscular fatty acid composition in an iberian x landrace cross. *J Anim Sci* 90: 2883-2893.
- Ramos, A. M., R. P. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, P. Dehais, M. S. Hansen, J. Hedegaard, Z. L. Hu, H. H. Kerstens, A. S. Law, H. J. Megens, D. Milan, D. J. Nonneman, G. A. Rohrer, M. F. Rothschild, T. P. Smith, R. D. Schnabel, C. P. Van Tassell, J. F. Taylor, R. T. Wiedmann, L. B. Schook, and M. A. Groenen. 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.
- Rankin, E. B., J. Rha, M. A. Selak, T. L. Unger, B. Keith, Q. Liu, and V. H. Haase. 2009a. Hypoxia-inducible factor 2 regulates hepatic lipid metabolism. *Molecular and Cellular Biology* 29: 4527-4538.

- Rankin, E. B., J. Rha, M. A. Selak, T. L. Unger, B. Keith, Q. Liu, and V. H. Haase. 2009b. Hypoxia-inducible factor 2 regulates hepatic lipid metabolism. *Mol Cell Biol* 29: 4527-4538.
- Ravasi, T., H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C. O. Daub, A. R. R. Forrest, J. Gough, S. Grimmond, J.-H. Han, T. Hashimoto, W. Hide, O. Hofmann, A. Kamburov, M. Kaur, H. Kawaji, A. Kubosaki, T. Lassmann, E. van Nimwegen, C. R. MacPherson, C. Ogawa, A. Radovanovic, A. Schwartz, R. D. Teasdale, J. Tegnér, B. Lenhard, S. A. Teichmann, T. Arakawa, N. Ninomiya, K. Murakami, M. Tagami, S. Fukuda, K. Imamura, C. Kai, R. Ishihara, Y. Kitazume, J. Kawai, D. A. Hume, T. Ideker, and Y. Hayashizaki. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 140: 744-752.
- Reverter, A., and E. K. Chan. 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24: 2491-2497.
- Schug, T. T., and X. Li. 2011. Sirtuin 1 in lipid metabolism and obesity. *Annals of Medicine* 43: 198-211.
- Seki, Y., L. Williams, P. M. Vuguin, and M. J. Charron. 2012. Minireview: Epigenetic programming of diabetes and obesity: Animal models. *Endocrinology* 153: 1031-1038.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for

- integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504.
- Snel, B., G. Lehmann, P. Bork, and M. A. Huynen. 2000. String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research* 28: 3442-3444.
- Szklarczyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguéz, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. v. Mering. 2011. The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* 39: D561-D568.
- Takahashi, N., T. Kawada, T. Yamamoto, T. Goto, A. Taimatsu, N. Aoki, H. Kawasaki, K. Taira, K. K. Yokoyama, Y. Kamei, and T. Fushiki. 2002. Overexpression and ribozyme-mediated targeting of transcriptional coactivators creb-binding protein and p300 revealed their indispensable roles in adipocyte differentiation through the regulation of peroxisome proliferator-activated receptor γ . *Journal of Biological Chemistry* 277: 16906-16912.
- Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, C. T. Johansen, S. W. Fouchier, A. Isaacs, G. M. Peloso, M. Barbalic, S. L. Ricketts, J. C. Bis, Y. S. Aulchenko, G. Thorleifsson, M. F. Feitosa, J. Chambers, M. Orho-Melander, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho, M. Jin Go, Y. Jin Kim, J.-Y. Lee, T. Park, K. Kim, X. Sim, R. Tzee-Hee Ong, D. C. Croteau-Chonka, L. A. Lange, J. D. Smith, K. Song, J. Hua Zhao, X. Yuan, J. a. Luan, C. Lamina, A. Ziegler, W. Zhang, R. Y. L. Zee, A. F. Wright, J. C. M. Witteman, J.

F. Wilson, G. Willemsen, H. E. Wichmann, J. B. Whitfield, D. M. Waterworth, N. J. Wareham, G. Waeber, P. Vollenweider, B. F. Voight, V. Vitart, A. G. Uitterlinden, M. Uda, J. Tuomilehto, J. R. Thompson, T. Tanaka, I. Surakka, H. M. Stringham, T. D. Spector, N. Soranzo, J. H. Smit, J. Sinisalo, K. Silander, E. J. G. Sijbrands, A. Scuteri, J. Scott, D. Schlessinger, S. Sanna, V. Salomaa, J. Saharinen, C. Sabatti, A. Ruukonen, I. Rudan, L. M. Rose, R. Roberts, M. Rieder, B. M. Psaty, P. P. Pramstaller, I. Pichler, M. Perola, B. W. J. H. Penninx, N. L. Pedersen, C. Pattaro, A. N. Parker, G. Pare, B. A. Oostra, C. J. O'Donnell, M. S. Nieminen, D. A. Nickerson, G. W. Montgomery, T. Meitinger, R. McPherson, M. I. McCarthy, W. McArdle, D. Masson, N. G. Martin, F. Marroni, M. Mangino, P. K. E. Magnusson, G. Lucas, R. Luben, R. J. F. Loos, M.-L. Lokki, G. Lettre, C. Langenberg, L. J. Launer, E. G. Lakatta, R. Laaksonen, K. O. Kyvik, F. Kronenberg, I. R. Konig, K.-T. Khaw, J. Kaprio, L. M. Kaplan, A. Johansson, M.-R. Jarvelin, A. Cecile J. W. Janssens, E. Ingelsson, W. Igl, G. Kees Hovingh, J.-J. Hottenga, A. Hofman, A. A. Hicks, C. Hengstenberg, I. M. Heid, C. Hayward, A. S. Havulinna, N. D. Hastie, T. B. Harris, T. Haritunians, A. S. Hall, U. Gyllensten, C. Guiducci, L. C. Groop, E. Gonzalez, C. Gieger, N. B. Freimer, L. Ferrucci, J. Erdmann, P. Elliott, K. G. Ejebe, A. Doring, A. F. Dominiczak, S. Demissie, P. Deloukas, E. J. C. de Geus, U. de Faire, G. Crawford, F. S. Collins, Y.-d. I. Chen, M. J. Caulfield, H. Campbell, N. P. Burtt, L. L. Bonnycastle, D. I. Boomsma, S. M. Boekholdt, R. N. Bergman, I. Barroso, S. Bandinelli, C. M. Ballantyne, T. L. Assimes, T. Quertermous, D. Altshuler, M. Seielstad, T. Y. Wong, E. S. Tai, A. B. Feranil, C. W. Kuzawa, L. S. Adair, H. A. Taylor Jr, I. B.

- Borecki, S. B. Gabriel, J. G. Wilson, H. Holm, U. Thorsteinsdottir, V. Gudnason, R. M. Krauss, K. L. Mohlke, J. M. Ordovas, P. B. Munroe, J. S. Kooner, A. R. Tall, R. A. Hegele, J. J. P. Kastelein, E. E. Schadt, J. I. Rotter, E. Boerwinkle, D. P. Strachan, V. Mooser, K. Stefansson, M. P. Reilly, N. J. Samani, H. Schunkert, L. A. Cupples, M. S. Sandhu, P. M. Ridker, D. J. Rader, C. M. van Duijn, L. Peltonen, G. R. Abecasis, M. Boehnke, and S. Kathiresan. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707-713.
- Torres-Padilla, M. E., and M. C. Weiss. 2003. Effects of interactions of hepatocyte nuclear factor 4 α isoforms with coactivators and corepressors are promoter-specific. *FEBS Letters* 539: 19-23.
- Ulven, S. M., K. T. Dalen, J.-Ö. Gustafsson, and H. I. Nebb. 2005. Lxr is crucial in lipid metabolism. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 73: 59-63.
- Urbanucci, A., K. Waltering, H. Suikki, M. Helenius, and T. Visakorpi. 2008. Androgen regulation of the androgen receptor coregulators. *BMC Cancer* 8: 219.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.
- Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. 2009. A census of human transcription factors: Function, expression and evolution. *Nat Rev Genet* 10: 252-263.
- Wakil, S. J., and L. A. Abu-Elheiga. 2009. Fatty acid metabolism: Target for metabolic syndrome. *Journal of Lipid Research* 50: S138-S143.

- Wang, L., W. Zheng, H. Zhao, and M. Deng. 2013. Statistical analysis reveals co-expression patterns of many pairs of genes in yeast are jointly regulated by interacting loci. *PLoS Genet* 9: e1003414.
- Wang, X., J. Chen, H. Liu, Y. Xu, X. Wang, C. Xue, D. Yu, and Z. Jiang. 2008. The pig p160 co-activator family: Full length cdna cloning, expression and effects on intramuscular fat content in longissimus dorsi muscle. *Domestic Animal Endocrinology* 35: 208-216.
- Wang, X. L., R. Suzuki, K. Lee, T. Tran, J. E. Gunton, A. K. Saha, M.-E. Patti, A. Goldfine, N. B. Ruderman, F. J. Gonzalez, and C. R. Kahn. 2009. Ablation of *arnt/hif1 α* in liver alters gluconeogenesis, lipogenic gene expression, and serum ketones. *Cell Metabolism* 9: 428-439.
- Willert, K., J. D. Brown, E. Danenberg, A. W. Duncan, I. L. Weissman, T. Reya, J. R. Yates, and R. Nusse. 2003. Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* 423: 448-452.
- Wood, J. D., and M. Enser. 1997. Factors influencing fatty acids in meat and the role of antioxidants in improving meat quality. *Br J Nutr* 78 Suppl 1: S49-60.
- Wood, J. D., M. Enser, A. V. Fisher, G. R. Nute, R. I. Richardson, and P. R. Sheard. 1999. Manipulating meat quality and composition. *Proc Nutr Soc* 58: 363-370.
- Wood, J. D., M. Enser, A. V. Fisher, G. R. Nute, P. R. Sheard, R. I. Richardson, S. I. Hughes, and F. M. Whittington. 2008. Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci* 78: 343-358.

- Xie, X., H. Liao, H. Dang, W. Pang, Y. Guan, X. Wang, J. Y.-J. Shyy, Y. Zhu, and F. M. Sladek. 2009. Down-regulation of hepatic hnf4 α gene expression during hyperinsulinemia via srebps. *Molecular Endocrinology* 23: 434-443.
- Yang, F., X. Li, M. Sharma, C. Y. Sasaki, D. L. Longo, B. Lim, and Z. Sun. 2002. Linking β -catenin to androgen-signaling pathway. *Journal of Biological Chemistry* 277: 11336-11344.
- Zhang, Y., K. Klein, A. Sugathan, N. Nassery, A. Dombkowski, U. M. Zanger, and D. J. Waxman. 2011. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. *PLoS ONE* 6: e23506.
- Zhou, D., R. S. Strakovsky, X. Zhang, and Y. X. Pan. 2012. The skeletal muscle wnt pathway may modulate insulin resistance and muscle development in a diet-induced obese rat model. *Obesity (Silver Spring)* 20: 1577-1584.