Universitat Autònoma de Barcelona

FUNCTIONAL ANALYSIS OF CANDIDATE GENES FOR MEAT QUALITY TRAITS AND MUSCLE TRANSCRIPTOMICS IN PIGS

Lourdes Criado Mesas

UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Ciència Animal i dels Aliments, Facultat de Veterinària

CENTRE DE RECERCA EN AGRIGENÒMICA

Departament de Genètica Animal

FUNCTIONAL ANALYSIS OF CANDIDATE GENES FOR MEAT QUALITY TRAITS AND MUSCLE TRANSCRIPTOMICS IN PIGS

Lourdes Criado Mesas

Doctoral thesis to obtain the PhD degree in Animal Production of the Universitat Autònoma de Barcelona, February 2020

Supervisor

Dr. Josep Maria Folch Albareda

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

fa constar

que **Lourdes Criado Mesas** ha realitzat sota la seva direcció el treball de recerca "*Functional analysis of candidate genes for meat quality traits and muscle transcriptomics in pigs*"

i certifica

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

considerant

que la memòria resultant és apta per optar al grau de Doctora en Producció animal per la Universitat Autònoma de Barcelona,

i perquè quedi constància, signen aquest document a

Bellaterra, a 15 de febrer de 2020.

Dr. Josep Maria Folch Albareda Lourdes Criado Mesas

Cover designed by Joan Jené.

This work was funded by the *Ministerio de Economía y Competititvidad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) projects AGL2014-56369-C2-2-R and AGL2017-82641-R. We acknowledge the support of the Spanish *Ministerio de Economía y Competitividad* for the 'Severo Ochoa Programme for Centres of Excellence in R&D' 2016-2019 (SEV-2015-0533) grant awared to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

Lourdes Criado was financially supported by a *Formación del Personal Investigador* (FPI) grant from the AGL2014-56369-C2 project provided by the *Ministerio de Economía y Competititvidad* (MINECO).

“Dans la vie, rien n'est à craindre, tout est à comprendre.”

— Marie Curie

CONTENT

SUMMARY

Nowadays, global meat consumption is rising, being pork one of the most consumed meats. The improvement of meat quality is subjected to consumer preferences for healthier and tastier meat products. Fatty acid composition and intramuscular fat are important factors determining meat quality traits, although a complex network of processes and pathways are involved. In the present thesis we aimed to elucidate the molecular mechanisms affecting fatty acid composition in pigs.

We studied the expression of candidate genes affecting intramuscular fat deposition and fatty acid composition traits, selected in previous studies of our group. The mRNA expression level of 45 genes was measured by RT-qPCR in a total of 355 pigs belonging to three different backcrosses based on Iberian boards. The eGWAS identified two *cis*-eQTL regions for *IGF2* and *ACSM5* genes, and ten *trans*-eQTL effects. The eGWAS performed in each backcross individually revealed different eQTL regions as well as six *trans*-eQTL hotspots, two per backcross, suggesting that breed-specific genetic variants are regulating the expression of these candidate genes.

Furthermore, we aimed to study the association between the *IGF2:g.3072G>A* polymorphism and the *IGF2* gene expression, and its effect on fatty acid composition in backfat adipose tissue in 355 pigs belonging to the three backcrosses. The eGWAS identified a *cis*-eQTL region associated with the *IGF2* gene expression in adipose tissue, being the *IGF2:g.3072G>A* polymorphism the most significantly associated SNP. In addition, the *IGF2* gene expression in both muscle and adipose tissue was explained by an imprinting model. Finally, animals carrying the *A* allele showed a higher PUFA and lower MUFA percentages in adipose tissue.

To better understand the regulation and role of miR-33a and miR-33b in porcine lipid metabolism, we studied the expression of these miRNAs in liver, adipose tissue and muscle, and their association with fatty acid composition. A total of 42 pigs were analysed and different expression patterns among tissues were observed for both miRNAs, suggesting that expression regulatory mechanisms are tissue dependent. In adipose tissue and muscle, a high correlation between miR-33a and miR-33b expression was observed, indicating a similar regulation. Conversely, a low correlation between the

two miRNAs in liver suggests a different regulation and function, and the miR-33b may be involved in fatty acid β-oxidation. Lastly, significant correlations among the expression of miR-33a and miR-33b in liver and adipose tissue and the adipose tissue fatty acid composition reinforced the involvement of these miRNAs in the regulation of lipid metabolism.

Finally, a preliminary study on muscle transcriptome of 132 pigs by RNA-Seq was performed with the aim to identify potential muscle gene-expression regulators. The eGWAS identified a total of 324 eQTLs, of which 247 were classified as *cis*-eQTL and 77 as *trans*-eQTLs. The two most significant associations were found for *HGFAC* and *HUS1* genes. The main processes identified for the expression of 291 genes with significant eQTLs were metabolic pathways and the top three canonical pathways were Granzyme B signalling, glutathione-mediated detoxification and NRF2-mediated Oxidative Stress Response. At last, *HNF4A*, *KLF3, E2F4, mir-483* and *RORC* were proposed as the main transcription factors, nuclear receptors or miRNAs involved in the muscle gene expression regulation.

RESUMEN

Hoy en día el consumo mundial de carne está creciendo, siendo la carne de cerdo una de las más consumidas. La mejora de la calidad de la carne está sujeta a los requisitos del consumidor por productos cárnicos más saludables y sabrosos. La composición de los ácidos grasos y la grasa intramuscular son factores determinantes de la calidad de carne, aunque hay involucrada una red compleja de procesos y vías. El objetivo de la presente tesis fue dilucidar los mecanismos moleculares que afectan a la composición de ácidos grasos en cerdos.

Estudiamos la expresión de genes candidatos relacionados con los caracteres de deposición de la grasa intramuscular y la composición de ácidos grasos, seleccionados en estudios previos de nuestro grupo. Los niveles de expresión del ARNm de 45 genes se midieron mediante RT-qPCR en un total de 355 cerdos pertenecientes a tres retrocruces experimentales diferentes basados en machos Ibéricos. El eGWAS identificó dos regiones *cis*-eQTLs para los genes *IGF2* y *ACSM5*, y diez *trans*-eQTLs. El eGWAS realizado para cada retrocruce por separado reveló diferentes regiones eQTLs así como seis *trans*-eQTL *hotspots*, dos por retrocruce, sugiriendo que variantes genéticas específicas de cada raza están regulando la expresión de estos genes candidatos.

Además, nuestro objetivo fue estudiar la asociación entre el polimorfismo *IGF2:g.3072G>A* y la expresión del gen *IGF2*, así como su efecto en la composición de ácidos grasos de la grasa dorsal de 355 cerdos pertenecientes a los tres retrocruces. El eGWAS identificó una región *cis*-eQTL asociada a la expresión del gen *IGF2* en el tejido adiposo, siendo el polimorfismo *IGF2:g.3072G>A* el más significativamente asociado. Asimismo, la expresión del gen *IGF2*, tanto en músculo como en tejido adiposo, fue explicada por un modelo de impronta genética. Finalmente, los animales portadores del alelo A mostraron un mayor porcentaje de PUFA y un menor porcentaje de MUFA en tejido adiposo.

Para entender mejor la regulación y función del miR-33a y el miR-33b en el metabolismo de los lípidos en cerdo, estudiamos la expresión de estos miRNAs en hígado, tejido adiposo y músculo, y su asociación con la composición de los ácidos grasos. Un total de 42 cerdos fueron analizados y se observaron diferentes patrones de expresión para

ambos miRNAs según el tejido, lo que sugiere que los mecanismos reguladores de la expresión dependen del tejido. Las altas correlaciones entre la expresión del miR-33a y el miR-33b en tejido adiposo y músculo indicaron una regulación similar. Por el contrario, la baja correlación entre los dos miRNAs en hígado sugirió una función y regulación diferentes, y el miR-33b parece estar involucrado en la β-oxidación de los ácidos grasos. Por último, las correlaciones significativas entre la expresión del miR-33a y el miR-33b en hígado y tejido adiposo y la composición de los ácidos grasos del tejido adiposo refuerzan la hipótesis del papel de estos miRNAs en la regulación del metabolismo lipídico.

Finalmente, se realizó un estudio preliminar del transcriptoma del músculo en 132 cerdos del retrocruce BC1_DU mediante RNA-Seq, con el objetivo de identificar potenciales reguladores de la expresión génica en músculo. Los eGWAS identificaron un total de 324 eQTL, de los cuales 247 fueron clasificados como *cis*-eQTLs y 77 como *trans*-eQTLs. Las dos asociaciones más significativas se encontraron para los genes *HGFAC* y *HUS1*. Los principales procesos identificados para la expresión de los 291 genes con eQTLs significativos fueron las vías metabólicas y las tres principales vías canónicas fueron la señalización de la granzima B, la desintoxicación mediada por glutatión y la respuesta al estrés oxidativo mediada por NRF2. Por último, *HNF4A*, *KLF3*, *E2F4*, *mir-483* y el gen *RORC* se propusieron como los principales factores de transcripción, receptores nucleares o miRNAs implicados en la regulación de la expresión génica del músculo.

LIST OF TABLES

GENERAL INTRODUCTION

PAPER I

PAPER II

PAPER III

PAPER IV

GENERAL DISCUSSION

ANNEXES: Paper I

ANNEXES: Paper II

ANNEXES: Paper III

LIST OF FIGURES

GENERAL INTRODUCTION

PAPER I

PAPER II

PAPER III

PAPER IV

GENERAL DISCUSSION

ANNEXES: Paper II

ANNEXES: Additional study

LIST OF PUBLICATIONS

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

- Paper II. **Criado-Mesas L.**, Ballester, M., Crespo-Piazuelo, D., Castelló, A., Fernández, AI. and Folch, JM. (2019) 'Identification of eQTLs associated with lipid metabolism in Longissimus dorsi muscle of pigs with different genetic backgrounds. Scientific Reports (in revision).

- Paper I. Criado-Mesas L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Benítez, R., Fernández, AI. and Folch, JM. (2019) 'Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition'. PLoS One, 8;14(8):e0220708.

- Paper III. **Criado-Mesas L.**, Ballester, M., Crespo-Piazuelo, D., Passols, M., Castelló, A. and Folch, JM. (2019). 'Expression analysis of porcine miR-33a/b in liver, adipose tissue and longissimus dorsi muscle and its role in fatty acid metabolism'. Manuscript in preparation.

and the preliminary manuscript:

- Paper IV. **Criado-Mesas L.** *et al*. Unraveling porcine muscle gene expression regulators through expression genome-wide association studies.

RELATED PUBLICATIONS BY THE AUTHOR

(Not included in the thesis)

- Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., **Criado-Mesas, L.**, Castelló, A., Fernández, AI., Ballester, M., Folch, JM. (2018) 'Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions'. Scientific Reports, 8(1):2045.

- Crespo-Piazuelo, D., Estellé, J., Revilla, M., **Criado-Mesas, L.**, Ramayo-Caldas, Y., Óvilo, C., Fernández, AI., Ballester, M., Folch, JM. (2018) 'Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions'. Scientific Reports, 8(1):12727.

- Crespo-Piazuelo, D., Migura-Garcia, L., Estellé, J., **Criado-Mesas, L.**, Revilla, M., Castelló, A., Muñoz, M., García-Casco, JM., Fernández, AI., Ballester, M., Folch, JM. (2019) 'Association between the pig genome and its gut microbiota composition'. Scientific Reports, 9(1):8791

- Crespo-Piazuelo, D., **Criado-Mesas, L.**, Revilla M., Castelló, A., Fernández, AI., Folch, JM., Ballester, M. (2019) 'Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs'. Plos One, 14(6):e0218862

- Reyes-Camacho, D., Vinyeta, E., Pérez, JF., Aumiller, T, **Criado-Mesas, L.**, Palade, LM., Taranu, I., Folch, JM., Calvo, MA., Van der Klis, JD., Solà-Oriol, D. (2020) 'Phytogenic actives supplemented in hyperprolific sows: effects on maternal transfer of phytogenic compounds, colostrum and milk features, performance and antioxidant status of sows and their offspring, and piglet intestinal gene expression. *Journal of Animal Science*, 98(1): skz390.

ABBREVIATIONS

3BCs	The three backcrosses together
3C	Chromosome conformation capture
4C	Circularized chromosome conformation capture
ACACA	Acetyl-CoA carboxylase
ACSM5	Acyl-CoA synthethase medium-chain family member 5
BAC	Bacterial artificial chromosome
BC1_DU	25% Iberian x 75% Duroc backcross
BC1_LD	25% Iberian x 75% Landrace backcross
BC1_PI	25% Iberian x 75% Pietrain backcross
cDNA	Complementary DNA
ChIP-Seq	Chromatine Immunoprecipitation sequencing
CPT	Carnitine palmitoyltransferases
dsDNA	Double stranded DNA
ELOVL6	ELOVL fatty acid elongase 6
eQTL	Expression quantitative trait loci
FASN	FA synthase
GWAS	Genome-wide association study
IGF2	Insulin like growth factor 2
lncRNA	Long non-coding RNAs
miRNA	microRNA
MUFA	Monounsaturated fatty acid
NGS	Next generation sequencing
PCR	Polymerase-chain reaction
PPAR	Peroxisome proliferator activated receptor
PUFA	Polyunsaturated fatty acid
QTL	Quantitative trait loci
RNA-Seq	RNA sequencing
RT-qPCR	High-throughput real-time quantitative PCR
SFA	Saturated fatty acid

sncRNA	Small non-coding RNAs
SNP	Single nucleotide polymorphism
SREBF	Sterol regulatory element binding transcription factor
SSC	*Sus scrofa* chromosome
TGS	Third-generation sequencing

General Introduction

Chapter 1

1.1. Current situation of porcine meat production

The pig (*Sus scrofa*) is a species of interest in livestock due to its economic contribution, being one of the most consumed meats in the world together with chicken and beef. Pig was one of the first species to be domesticated, but the systematic improvement of its production started in 1960s and 1970s by crossing breeds and the establishment of selection programs.

Europe produced around 63.8 million tonnes of meat in 2018 and about one half (29.7 million tonnes) was from pigs, which has increased every year. Worldwide main producers of pork are Asia (57.2%), Europe (19.3%) and America (18.8%) (Figure 1.1.A.) (FAOSTAT, 2018).

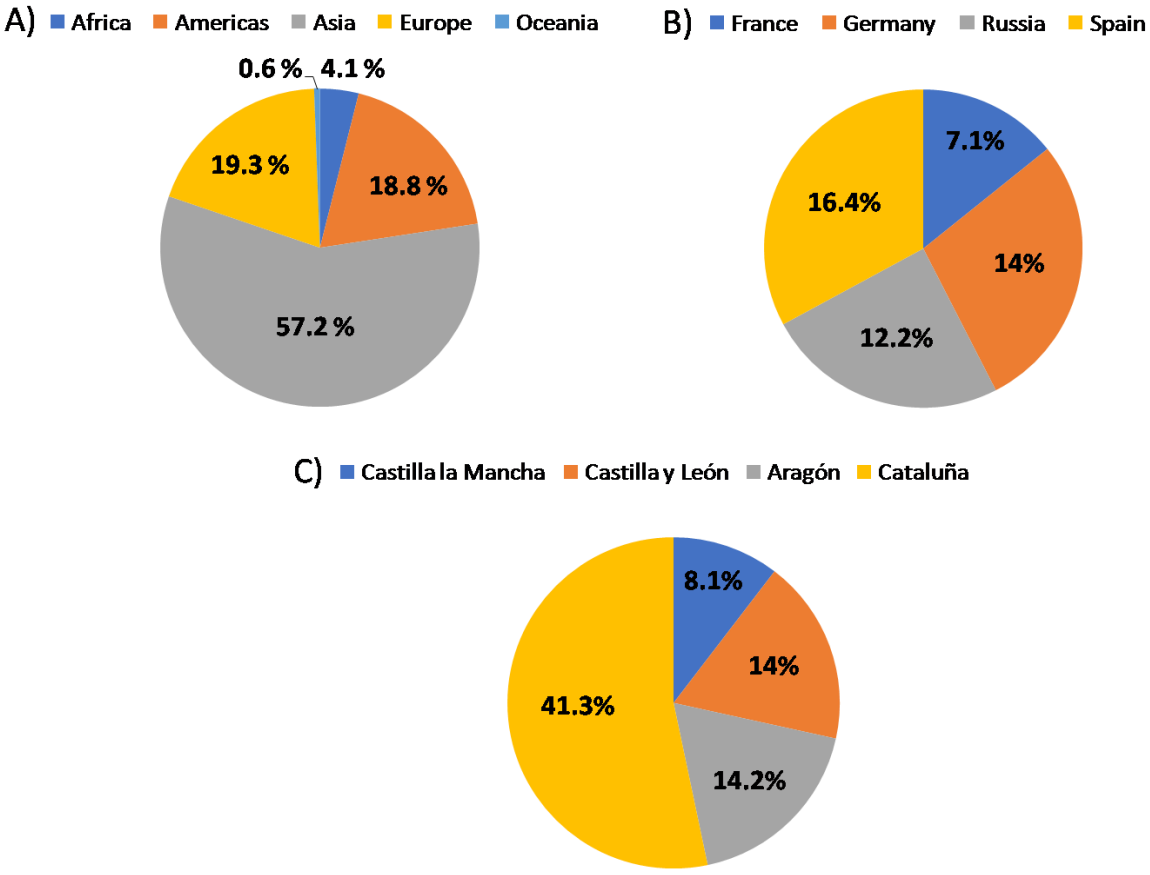


Figure 1.1. Percentage of world pig production in A) the world, B) in Europe (FAOSTAT, 2018), and C) in Spain (MAPA, 2018).

In Europe, the two main producers are Spain (16.4% of the Europe's pig meat production) and Germany (14% of the Europe's pig meat production), followed by Russia and France (12.2% and 7.1% respectively) (Figure 1.1.B and Figure 1.2.).

Finally, Catalonia is leading the Spanish pig production with 41.3% in 2018 (MAPA: Ministerio de Agricultura, Pesca y Alimentación, Gobierno de España, Junio 2019) followed by Aragón and Castilla León with 14.2% and 14%, respectively (Figure 1.1.C).

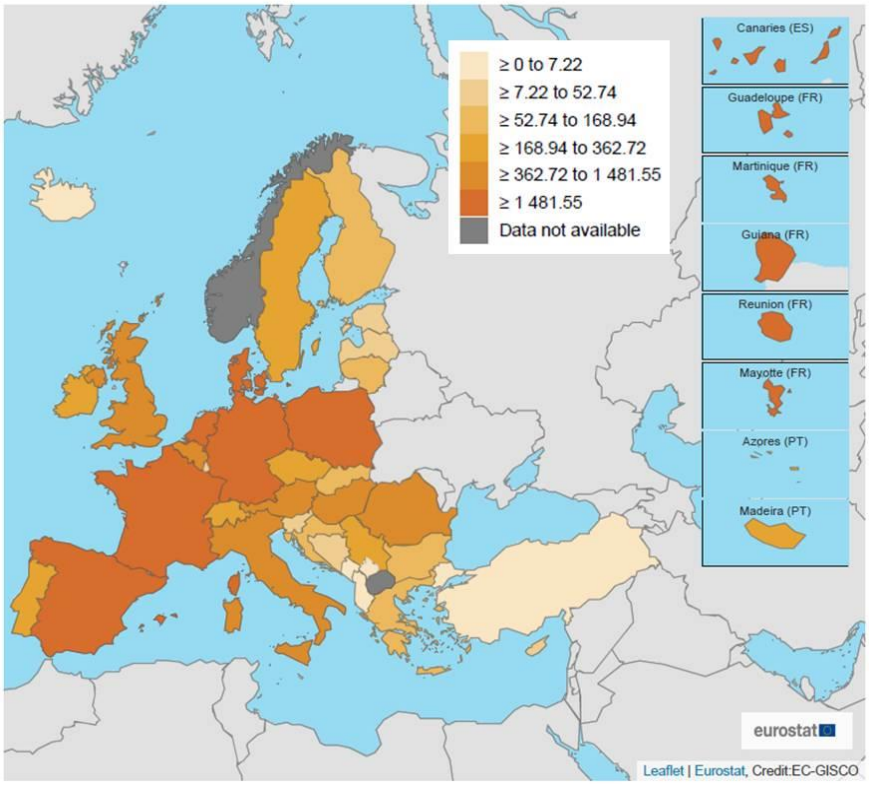


Figure 1.2. Average number of porcine meat production (thousands of tonnes) by region in Europe [Eurostat 2018; https://ec.europa.eu/eurostat/; accessed September 2019].

Global meat consumption is rising and differs among societies because it is directly related to people's incomes and population growth. Important economic, sanitary and environmental consequences are driven by this increase in meat consumption. In recent years there have been changes in the type and quantity of meat we eat, for example chicken and pork are gaining more interest, and now we eat more processed products. Hence, a high meat consumption and meat products have a considerable

effect on people's health and have been highly related with many diseases, for example colorectal cancer. Moreover, the increase in livestock production has an important negative impact on the environment because it is a major source of greenhouses gases, uses more fresh water than other human activities and produces soil erosion and changes in the biodiversity (Godfray *et al.*, 2018).

1.2. Pork meat quality traits

Meat quality traits are complex phenotypes difficult to measure because are subjected to different stakeholders, that is, producers, slaughterers, processors, distributors and consumers, with different requirements about quality traits, depending on the use of the product. Pork meat should be efficiently produced and with a required level of quality. Meat quality can be determined by several factors (Figure 1.3.) which include animal welfare (good ethical production practices), food safety (microbiological hazards), technological factors (pH, firmness, water-holding capacity, and cooking), sensorial aspects (aroma, texture, flavour, taste, juiciness, colour, and marbling), healthiness and nutritional values (intramuscular fat content, lipid composition, and digestibility), and serviceability (ease of use, ability to be processed, and prices) (Wood and Whittemore, 2007; Wood *et al.*, 2008; Listrat *et al.*, 2016).

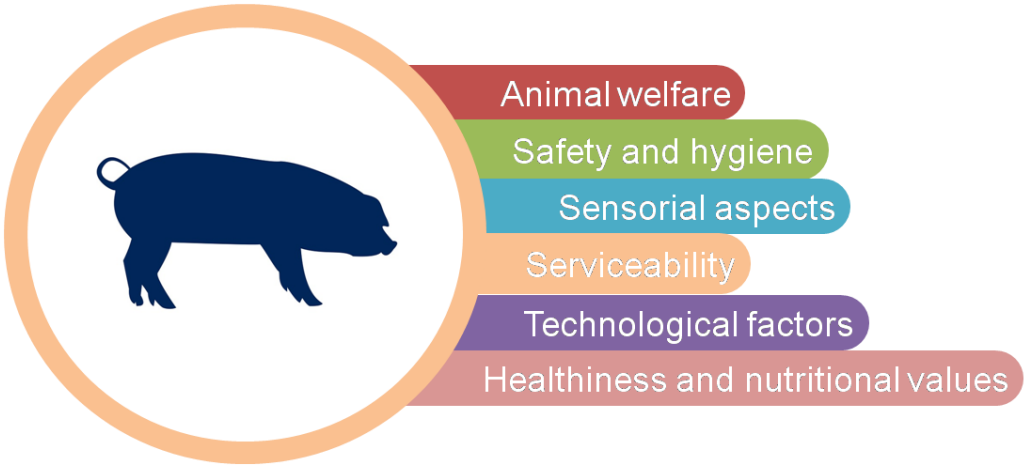


Figure 1.3. Main factors related with pork meat quality.

During the last years, consumer's requirements have changed and now taste and nutritional values are the two most important qualities attributes of meat. By one hand, tasty eating products are driven by the colour, texture, juiciness and flavour (Wood *et al.*, 2004; Wood and Whittemore, 2007). On the other hand, nutritional values are determined by fats, carbohydrates and proteins. It is accepted that the amount of fat in meat can influence tenderness and flavour, but most consumers consider fat as unhealthy because its relationship with modern life diseases, such as colorectal cancer or cardiovascular diseases (Wood *et al.*, 1999; Webb and O'Neill, 2008).

The main porcine breeding interests are growth, carcass quality, fertility, fatness, feed efficiency, disease resistance, behaviour, and meat quality. During years, a strong selection process was based on increasing the percentage of lean meat in the carcass due to its economic value, leading to a reduction of intramuscular fat in some breeds, negatively affecting meat quality caused by a reduction in taste and tenderness. Therefore, commercial animal breeds with a reduced backfat and high growth rate (e.g. Landrace) were generated. Hence, pig breeding programs have included meat quality to satisfy the increasing consumer demand for healthier and tastier meat products (Wood and Whittemore, 2007).

1.2.1. Intramuscular fat content

Intramuscular fat can be defined as the amount of fat measured within muscles, and it is composed by a sum of phospholipids, triglycerides and cholesterol. Intramuscular fat content differs according to gender, age, feeding, breed, and muscle types. Intramuscular fat variability depends on the number and size of intramuscular adipocytes (Hocquette *et al.*, 2010). A moderate-high heretability for intramuscular fat content, ranging from 0.39 to 0.59, was reported in different studies (Cameron, 1990; Suzuki *et al.*, 2005; Won *et al.*, 2018).

Intramuscular fat plays a key role in several meat quality traits, and meat with a high intramuscular fat content is considered of good quality because it gives flavour, juiciness, tenderness, and/or firmness to the meat. In addition, it affects the palatability and nutritional values of meat (Wood *et al.*, 2004).

1.2.2. Fatty acid composition

Fatty acids are the main type of lipids and are classified in three categories based on the number of double bonds: i) saturated fatty acids (SFAs) with no double bonds; ii) monounsaturated fatty acids (MUFAs) with one double bond; and iii) polyunsaturated fatty acids (PUFAs) with two or more double bonds. The ratio between unsaturated and SFAs is important for the healthiness and nutritional values of the product (Wood and Whittemore, 2007; Webb and O'Neill, 2008). In addition, the degree of fatty acid saturation and number of carbons that forms the fatty acid chain influences fat firmness and oiliness, changing the melting point. SFAs consumption has been associated with an increase of cholesterol and low-density lipoprotein blood levels and therefore with modern human diseases like obesity, cancer and cardiovascular diseases. MUFAs improve meat flavour and contribute to a better taste and lower oxidation rate of meat. On the contrary, PUFAs are more susceptible to be oxidized, which produces rancidity and a consequent reduction of meat quality. Both MUFAs and PUFAs, particularly omega-3 PUFAs, are involved in the reduction of total cholesterol concentration (Wood and Enser, 1997; Webb and O'Neill, 2008). Therefore, fatty acids in both muscle and adipose tissue are determinant of meat quality and its nutritional values (Wood *et al.*, 2008).

1.3. Fatty acid metabolism

Lipids are one of the major classes of biomolecules and have different functions. They are the major source of energy storage and are important for cell membrane structure and to establish cellular communications, for example as lipokines in the regulation of fatty acid metabolism (Cao *et al.*, 2008). Depending on the nutritional status, the fatty acid metabolism is altered and is affected by two main reactions: lipolysis or fatty acid β-oxidation and lipogenesis or *de novo* fatty acid synthesis. Whenever the body enters in the fasting state, stored triglycerides are broken and lipolysis occurs, providing energy for the cells (Frühbeck *et al.*, 2014). On the contrary, during the fed state lipogenesis occurs, and carbohydrates are converted to fatty acids and stored as

triglycerides, an energy reservoir. Therefore, fatty acids can be provided by the diet or synthesized endogenously via *de novo* lipogenesis (Ameer *et al.*, 2014).

1.3.1. Fatty acid β-oxidation

During periods of decreasing food intake, prolonged fasted state or increasing energy demand, fatty acids are used from the adipose tissue storage to produce energy through mitochondrial fatty acid β-oxidation. This energy is produced by generating reducing agents ($FADH_2$ and $NADH^+$) to serve as electron donors to the respiratory chain for oxidative phosphorylation and ATP generation. Breakdown of fatty acid up to 18 carbons occurs directly in the mitochondria while longer fatty acids are first shortened in peroxisomes and then oxidized in the mitochondria (Eaton, Bartlett and Pourfarzam, 1996).

The fatty acid β-oxidation pathway (Figure 1.4.) takes place by several reactions catalyzed by different enzymes and all act on CoA esters. A preliminary step is the ATP-dependent formation of fatty acyl-CoA esters from free fatty acids, catalysed by acyl-CoA synthase, which are then introduced into the mitochondria through the carnitine palmitoyltransferases (CPT) system. CPT1 transfer the acyl-CoA through the inner mitochondrial membrane as carnitine and then the opposite reaction was done with CPT2, separating the carnitine from the acyl-CoA ester. Once inside the mitochondria, a few four-step cycles (dehydrogenation, hydratation, second dehydrogenation and thiolysis) were performed and a fatty acyl-CoA is shortened by two carbons to obtain acetyl-CoA, NADH and $FADH_2$ molecules. In the case of PUFAs, β-oxidation occurs at low rates and PUFA-CoA can act as a fatty acid β-oxidation inhibitor because they can contain *cis* double bonds at even-numbered carbon atoms (Eaton, Bartlett and Pourfarzam, 1996).

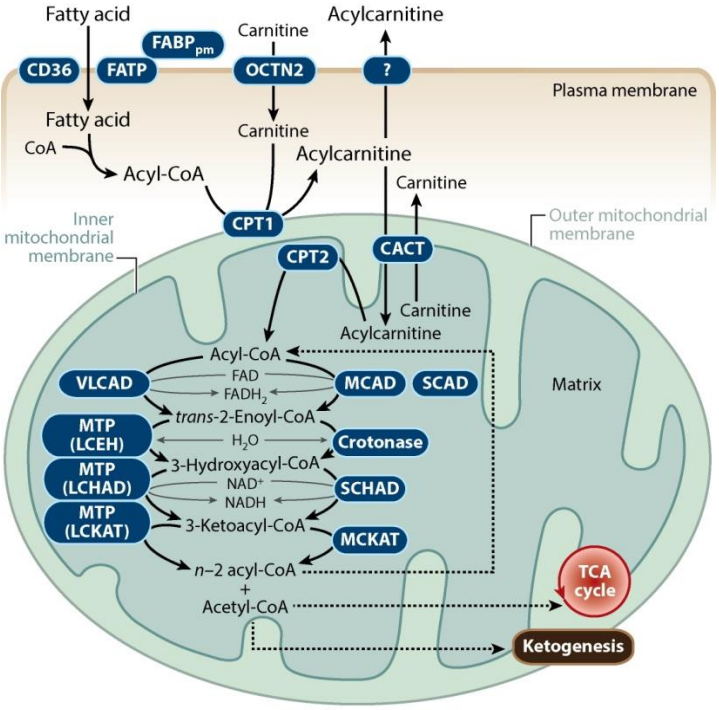


Figure 1.4. Schematic representation of mitochondrial fatty acid β-oxidation (reprinted from Houten *et al.*, 2016).

1.3.2. *De novo* fatty acid synthesis

Lipogenesis is a metabolic pathway involved in the synthesis of fatty acids from excess carbohydrates and then can be incorporated into triglycerides for energy storage. Dietary fatty acids are digested on the stomach, where lipids are partially digested by gastric lipases and then are moved into the intestinal track, where they are hydrolyzed by the pancreatic lipase producing monoacylglicerol, diacylclycerol and free fatty acids. In the enterocytes, fatty acids and monoacylglycerol are absorbed and re-esterificated to form triacylglycerol. Finally, chylomicrons formed from triacylglycerol together with cholesterol, phospholipids, and proteins are transported and incorporated into tissue lipids where they are metabolized (Wood *et al.*, 1999; Ameer *et al.*, 2014). The three main metabolic tissues are adipose tissue, liver and skeletal muscle, and they cooperate to supply energy requirements. In pigs, adipose tissue is the primary site for *de novo* fatty acid synthesis (O'Hea and Leveille, 1969), meanwhile in other species such as humans or rodents, liver is the target tissue. In addition, glucose is the main source of acetyl-CoA in pigs meanwhile acetate is used in other species due their poor

glucose metabolism. In general (Figure 1.5.), glucose enters to the glycolytic pathway and produces pyruvate, which is converted into acetyl-CoA that feeds the tricarboxylic acid cycle and produces citrate, which is converted back into acetyl-CoA. The acetyl-CoA obtained is carboxylated by the acetyl-CoA carboxylase (ACACA) enzyme to generate malonyl-CoA, which is then used as substrate for the production of palmitic acid (C16:0) by the fatty acid synthase (FASN) enzyme (Bergen and Mersmann, 2005; Ameer *et al.*, 2014). Subsequently, fatty acids taken from the diet and *de novo* synthesized fatty acids suffer different cycles of elongations and desaturations. Fatty acid elongation involves the addition of two-carbons of the malonyl-CoA group, while each acyl-CoA of a fatty acid can be desaturated by the introduction of a double bond in a specific position producing different MUFAs and PUFAs (Guillou *et al.*, 2004).

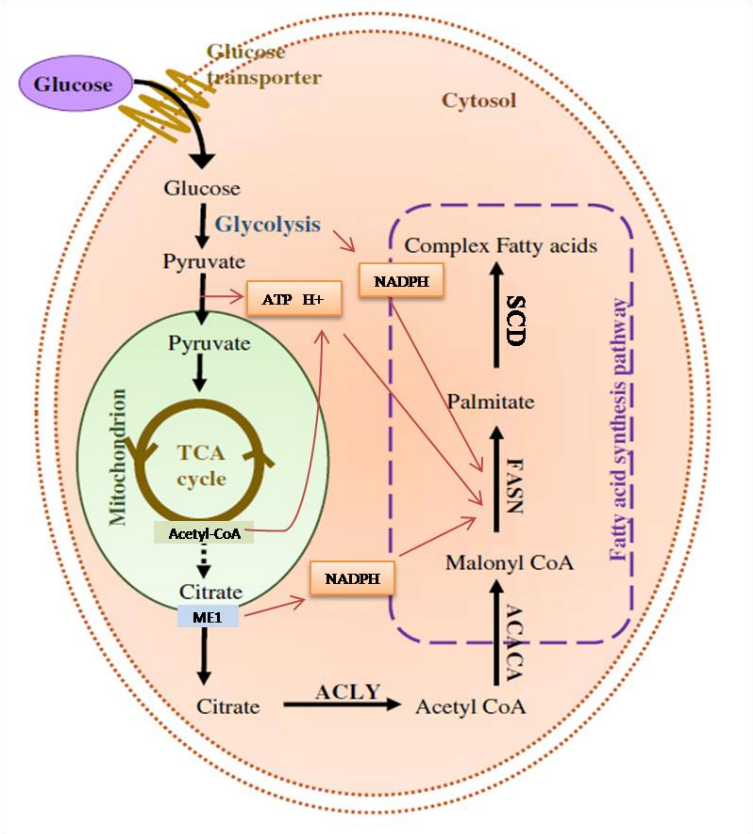


Figure 1.5.: Scheme of *de novo* lipogenesis (adapted from Ameer *et al.*, 2014).

1.4. Pig genomics

Over the past 60 years, advances in animal breeding, nutrition and management helped to improve efficiency in pork production. Animal breeding allows animal selection according to the genetic value that has been assigned to each one using different methods and genetic progress can be made by measurement of the interest traits (Wu and Bazer, 2019). In pigs, genetic evaluation approach has a strong impact on the improvement of the efficiency of pork production and on carcass quality. Genomic selection of relevant traits can be performed by increasing the accuracy of the prediction of the breeding values and by obtaining earlier evaluations. For instance, sow prolificacy traits tend to have low heritability and are only expressed in the mature females. Moreover, genomic selection can be valuable for traits which cannot be evaluated in breeding animals, such as meat quality traits.

In 2003, the Swine Genome Sequencing Consortium (SGSC) started the sequencing of the pig genome (Schook *et al.*, 2005) and in 2012 the *Sscrofa 10.2* assembly was published (Groenen *et al.*, 2012). Two different strategies were followed, first a hierarchical shotgun Sanger sequencing of bacterial artificial chromosome (BAC) clones (Humphray *et al.*, 2007), which was later supplemented with Illumina Next-generation sequencing (NGS) data obtained through whole-genome shotgun sequencing (Archibald *et al.*, 2010). The 2.596 Mb sequence of the *Sscrofa* 10.2 assembly was obtained from a single female Duroc animal. Nowadays hundreds of pigs of different breeds have been re-sequenced, and their genomes are public. In 2017, an improvement of the previous assembly was made and *Sscrofa* 11.1 assembly was available. This new assembly was constructed with data obtained through third-generation sequencing (TGS) technologies (PacBio RSII long reads), generating a 65x genome coverage over a total sequence length of 2.5 Gb.

NGS technologies allowed the massive detection of single nucleotide polymorphisms (SNPs) in pig genomes (Ramos *et al.*, 2009) and the development of high-throughput genotyping arrays, which consist of a collection of SNPs distributed along the entire pig genome. The first array, the *PorcineSNP60 BeadChip* (*Illumina*), was commercialized in 2008, before the completion of the pig genome sequence. This array contained 62,163

SNPs distributed along all pig chromosomes. Later, the *Axiom Porcine Genotyping Array* (*Affimetrix*) was commercialized and this array contains 658,692 markers, including 56,000 SNPs from the Illumina's chip, allowing compatibility with previous studies. Other low-density SNP panels have been developed in several studies with the purpose of reducing genotyping costs like the *GeneSeek Genomic Profiler for Porcine LD* (GeneSeek/Neogen), which contains 10,241 SNPs. The availability of these high-density panels covering the whole genome along with powerful statistical tools can provide significant insights into the molecular basis of phenotypic variation of production traits and assist breeders in pig selection, and are the base of genomic selection (Miar, Plastow and Wang, 2015).

Several 'omics' have been also applied to pigs (Table 1.1.). These new 'omics' technologies outline the system genetics approach, which integrates different levels of information like genomics (high-density genotyping and DNA sequencing), epigenomics (bisulfite sequencing, chromatine immunoprecipitation sequencing (ChIP-Seq), DNase-Sequencing, chromosome conformation capture (3C), circularized chromosome conformation capture (4C), and Hi-C), transcriptomics (microarrays, RNA-Sequencing (RNA-Seq) and high-throughput real-time quantitative polymerase-chain reaction (RT-qPCR)), proteomics (tandem mass spectrophotometry), metabolomics (gas chromatography and high-performance liquid chromatography-mass spectrophotometry), microbiomics (16S rRNA sequencing and whole-metagenome shotgun sequencing) and phenomics (image or video analysis-based) (MacKay, Stone and Ayroles, 2009; Ohashi *et al.*, 2015; Suravajhala, Kogelman and Kadarmideen, 2016).

Table 1.1. Description of the main 'omics' and their technologies

'Omics' level	Definition	Technology
Genomics	Analysis of the structure and function of a genome	Whole-genome sequencing
		Whole-exome sequencing
		High density genotyping
Epigenomics	Analysis of chemical modifications, chromatin structure, conformation, and its interaction with proteins	Bisulfite sequencing
		ChIP-Seq
		DNase-Seq
		3C and 4C
		HiC
Transcriptomics	Study of the expression levels of all gene transcripts in a particular cell, at a particular time, and in a particular state.	Microarrays
		RNA-Seq
		High throughput RT-qPCR
		Single-cell transcriptome analysis
Proteomics	Detection of quantitative and/or qualitative variation on proteins	Tandem mass spectrophotometry
Metabolomics	Detection of quantitative and/or qualitative variation on metabolites	Gas chromatography
		Mass spectrophotometry
		Nuclear magnetic resonance
Microbiomics	Study of the microbiota, their genomes and the surrounding environmental conditions from an entire habitat	16S rRNA sequencing
		Whole-metagenome shotgun sequencing
Phenomics	Collection of a high number of phenotypic data	Image or video analysis-based

1.4.1. Gene expression studies

Quantification of mRNA expression can be performed by different methodologies and the first known technique was the Northern blot hybridization which is based on the intensity of the hybridized band (Streit *et al.*, 2009). Later, other techniques such as serial analysis of gene expression, microarrays or high throughput sequencing allowed quantifying the expression of thousands of genes (Velculescu *et al.*, 1995; Edwin M. Southern, 2001).

Until de arrival of RNA-seq, microarrays were the standard method for gene expression quantification because were developed to analyze the expression of thousands of genes in a single reaction, although required sophisticated investments.

The first array commercialized was the Porcine AROS v1.0, *Operon*; Gene-Chip Porcine microarray (*Affymetrix*) in 2003 and consisted of a set of 10,665 oligo set. Latterly, these arrays were improved and customized and become a powerful tool for detecting differential gene expression. For instance, the GeneChip® Porcine Genome Array from Affymetrix contains 23,937 probe sets that interrogate approximately 20,201 *Sus scrofa* genes was the most widely used, but other arrays were commercialized such as the *PigOligoArray* from Illumina, which contains 20,400 70-mer oligonucleotides and the Snowball array from Affymetrix comprises 1,091,987 probes (47,845 probe sets) with a mean coverage of 22 probes/transcript (Steibel *et al.*, 2009; Freeman *et al.*, 2012). Microarrays in pigs were used to identify differentially expressed genes among the muscle transcriptome of animals with different intramuscular fat content and composition (Liu *et al.*, 2009; Cánovas *et al.*, 2010; D'Andrea *et al.*, 2011; Damon *et al.*, 2012; Hamill *et al.*, 2013; Pena *et al.*, 2013; Sun *et al.*, 2013; Yu *et al.*, 2013; González-Prendes *et al.*, 2019).

The microarray technology was progressively replaced by sequencing methods. Since the development of the dideoxy method of DNA sequencing by Sanger (Sanger, Nicklen and Coulson, 1977) a continuous improvement in the capacity and a reduction in cost have been achieved. In the 90s, the automatic sequencing using fluorescent terminators allowed the sequencing of the first genomes in humans and domestic animals. In the 2000s NGS methods were developed and commercialized allowing the

massive parallel sequencing of DNA. They have some advantages: i) do not require bacterial cloning of DNA fragments, instead libraries are prepared in a cell free system; ii) can run thousands-to-many-millions of sequencing reactions in parallel; iii) is not necessary an electrophoresis as the sequence output is directly detected; iv) a massive amount of sequences can be generated in a short period of time. However, the relative short reads obtained at the beginning was a disadvantage that make necessary the development of new alignment algorithms to perform the genome assembly (L.van Dijk *et al.*, 2014). NGS can be applied to whole genome sequence or to the sequencing of transcriptomes, which is called RNA-Seq. Several RNA-Seq studies have reported differentially expressed genes in pigs associated with sex, breed, growth and meat quality traits (Zhao *et al.*, 2011; Esteve-Codina *et al.*, 2011; Pérez-Montarelo *et al.*, 2012; Ramayo-Caldas *et al.*, 2012a; Corominas *et al.*, 2013a; Jiang *et al.*, 2013; Puig-Oliveras *et al.*, 2014; Sodhi *et al.*, 2014; Xing *et al.*, 2014; Ghosh *et al.*, 2015; Cardoso *et al.*, 2018). Both microarrays and RNA-Seq techniques are suitable for gene expression quantification, but microarrays are limited to the low sensibility and the high background. Hence, RNA-Seq allows to determine the transcript abundance with a larger dynamic range of expression levels (Table 1.2.) (Nookaew *et al.*, 2012). In addition, the QuantSeq 3' mRNA sequencing for RNA quantification was developed to sequence close to the 3' end of polyadenylated RNAs. This reduced substantially the price, as well as the sample preparation, sequencing and data processing in comparison to standard RNA-Seq (Moll *et al.*, 2014).

Table 1.2. Advantages and limitations of three transcriptomic methods.

Technique	Advantages	Limitations
Microarray	Low cost, large number of samples and high throughput	Limited number of genes, low sensitivity and high background
RNA-Seq	High accuracy and specificity, low background, high dynamic range and identification of novel transcripts, splice junctions, SNPs and non-coding RNAs	High cost, requires a NGS platform and high bioinformatics tools for data analysis and high data storage
RT-qPCR	Low cost, fast, high accuracy and specificity and wide dynamic range	Limited number of genes and requirement for specific primers

In the 1990s, the RT-qPCR was the preferred method for either single or multiple gene expression quantification. The basis of RT-qPCR is to monitor the process of DNA polymerase-chain reaction (PCR) in "real-time", meaning that can detect the amplification of the PCR amplicons at the end of each amplification cycle by using a fluorescent dye system and a thermocycler with fluorescence-detection capability. RT-qPCR is faster and less expensive compared to other RNA quantification methods, such as RNA-Seq, conferring an accurate and specific high-throughput mRNA quantification over a wide dynamic range (Table 1.2.) (Kuang *et al.*, 2018).

In RT-qPCR, the starting material could be total RNA, mRNA or other sources of RNA which is transcribed into complementary DNA (cDNA) by a reverse transcriptase enzyme and finally used as the material for PCR amplification. PCR primers should be designed in separate exons or spanning and exon-exon junction or RNA sample can be treated with RNAse free DNAse I or dsDNAse in order to avoid genomic DNA contamination. Once the cDNA is obtained it is amplified by PCR under optimized conditions to measure the expression level of target genes. Finally, data is analyzed using suitable normalization methods. The two main mRNA quantification strategies are: i) absolute quantification, which is relative to an external standard curve and

allows the generation of specific, sensitive and reproducible quantification data, or ii) relative quantification to one or more mRNAs from reference genes which expression does not change under the experimental conditions and must be carefully selected (Pfaffl, 2012). On the other hand, there are two principal methods according to the detection chemistries: i) a non-probe based chemistry, which is based on a fluorochrome that binds in a non-specific way to double stranded DNA (dsDNA), and the quantification is based on the exponential detection of the fluorescent signal (for example: SYBR Green) or ii) a probe-based chemistry that uses a fluorescent labelled oligonucleotide which hybridizes within the amplicon and aids to quantify changes in fluorescence only if the sequence is amplified (for example: Taqman probes) (Bustin and Nolan, 2004). All the RT-qPCR procedure (RNA extraction, integrity, cDNA synthesis, primer design, etc.) should follow the MIQE Guidelines (Bustin *et al.*, 2009) in order to be reproducible.

Several studies found differentially expressed genes in pig muscle in association with meat quality traits using RNA-Seq and were further validated by RT-qPCR (Gorni *et al.*, 2011; Óvilo *et al.*, 2014; Ayuso *et al.*, 2015; Li *et al.*, 2016; Muñoz *et al.*, 2018b; Piórkowska *et al.*, 2018; Gao *et al.*, 2019). In addition, differentially expressed genes for intramuscular fat were found and validated in adipose tissue (Xing *et al.*, 2019a; Zhao *et al.*, 2019). On the other hand, numerous single gene expression studies have been reported to study candidate genes for specific traits related to lipid metabolism: *ACSL4* (Corominas *et al.*, 2012), *APOA2* (Ballester *et al.*, 2016), *DGAT1* and *DGAT2* (Cui *et al.*, 2011), *ELOVL6* (Corominas *et al.*, 2015), *FABP4* and *FABP5* (Ballester *et al.*, 2017a), and *FADS2* (Gol *et al.*, 2018) among others. Later, array platforms appeared to study gene expression by multiplex RT-qPCR with customized designs. These new technologies are for example the Fluidigm Dynamic Array (*Fluidigm*) (Spurgeon, Jones and Ramakrishnan, 2008) or the Taqman Open Array platforms (*Life Technologies*) and allowed to study tens of genes in up to 96 animals per array in a cost-effective manner. In our group, a selected set of candidate genes for lipid metabolism in three different porcine tissues was quantified in a customized Fluidigm array (Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017b; Revilla *et al.*, 2018).

1.4.2. Regulation of gene expression

Gene expression can be controlled by several mechanisms acting mainly at two different levels, transcriptional and post-transcriptional, which are controlling gene expression from the transcription to the post-translational modifications (Figure 1.6.).

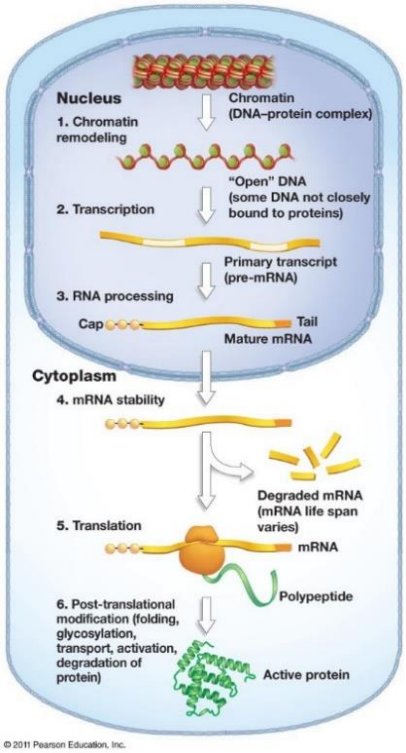


Figure 1.6. Overview of gene expression regulation at different levels.

Transcriptional regulation was considered the most important step in gene expression and it is easier to study with the established methods. At this level, regulation can be controlled by proteins that can be classified in two groups: sequence-specific DNA binding proteins (such as transcription factors) and proteins of large multiple-protein RNA polymerase machines (such as TATA-binding proteins). In addition, epigenetic mechanisms, such as DNA methylation and histone modifications, play an important role in transcriptional regulation.

Although transcriptional regulation was the most studied level, post-transcriptional regulation has emerged as relevant in many biological processes and it provides a more rapid response to cellular signals and/or environmental stimulus (Mata,

Marguerat and Bähler, 2005; López-Maury, Marguerat and Bähler, 2008). Small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs) are known factors in the post-transcriptional regulation. The sncRNAs class include small interfering RNAs, microRNAs (miRNAs), PIWI-interacting RNAs, endogenous small interfering RNAs, promoter associate RNAs, small nucleolar RNAs, and sno-derived RNAs, while lncRNAs includes long intervening/intergenic noncoding RNAs, natural antisense transcripts, enhancer RNAs, circular RNAs, competing endogenous RNAs, and promoter upstream transcripts (Filipowicz, Bhattacharyya and Sonenberg, 2008; Bergen and Burnett, 2013).

1.4.2.1. MicroRNAs

The first miRNA was discovered in 1993 by the Ambros and Ruvkun groups in *Caenorhabditis elegans* (Lee, Feinbaum and Ambros, 1993; Wightman, Ha and Ruvkun, 1993) and since then, miRNAs have been identified in different organisms, and some of them have been shown to be highly conserved across species. They play important roles in diverse regulatory pathways of many cellular processes and diseases, so the quantification of their expression can contribute to both diagnostic and prognostic of many diseases (Lagos-Quintana *et al.*, 2002; He and Hannon, 2004).

miRNAs are a class of small non-protein coding RNAs of 20-25 nucleotides long and are involved in post-transcriptional gene regulation by degrading the mRNA or by preventing the mRNA translation to protein. Most miRNAs are transcribed from DNA sequences into primary miRNAs and processed into precursor miRNAs and finally mature miRNAs. In most cases miRNAs interact with the 3'UTR of target miRNAs but they can also interact with other regions such as the 5'UTR, coding sequence and gene promoters (O'Brien *et al.*, 2018). The nucleotide sequence of the miRNA that specifically binds to the mRNA target side is called the "seed" region and it is located between positions 2 and 7 in 5'-3' direction. Members of the same miRNA family present a high homology in the seed region and miRNA binding sites are widely conserved in different species (Cai *et al.*, 2009). Their specific biological role remains unclear, but both functional characterization and miRNA target genes computer-based predictions suggested that miRNAs are involved in different cellular processes, such as

development, differentiation, proliferation and apoptosis (Kloosterman and Plasterk, 2006).

In the miRBase database, information of only 457 mature miRNAs in *Sus Scrofa* was available in comparison with 2,654 human miRNAs and 1,978 mouse miRNAs (Kozomara, Birgaoanu and Griffiths-Jones, 2019).

In 2005, miR-17, miR-18a, miR-19a, miR-20a, miR-19b and miR-92a were the first porcine miRNAs identified based on sequence homology with human miRNAs (Sawera *et al.*, 2005). Later, using deep sequencing and computational analysis several miRNAs were identified in different porcine tissues: muscle, fat, embryo, pituitary, intestine, ovary and testes (Song *et al.*, 2018). miRNAs that affect development and growth of skeletal muscle are gaining relevance due to the economic importance of the muscle traits. A list of miRNAs involved in myogenesis and muscle development was reviewed by Song *et al.* (2018), which provided insights into miRNA regulation of muscle growth and identified potential candidate genes for meat quality traits. miR-1, miR-133, and miR-206 were listed as the highest expressed miRNAs in porcine muscle (Song *et al.*, 2018). On the other hand, adipose tissue is also involved in meat quality and plays a key role in metabolic health. miR-143 was the first miRNA studied in adipose cell biology and miR-210 and miR-27 were described to be involved in adipogenesis in pigs (Wang, Gu and Jiang, 2013; Song *et al.*, 2018). As well, miR-215, miR-135, miR-224, miR-146b, miR-1a, miR-133a, miR-122, miR-204 and miR-183 were differentially expressed between two different breeds according to fat, so were described to be involved in adipose tissue development and growth in pigs (Wang, Gu and Jiang, 2013). Also, miR-33 was reported to play an important role in lipogenesis in the porcine adipose tissue, and target several genes related with lipid metabolism (Taniguchi *et al.*, 2014). Liver is a central organ which regulates lipid synthesis and metabolism in mammals, so many miRNA studies were also carried out in this metabolic tissue. For instance, the miR-122 was the first miRNA identified to regulate lipid metabolism in humans, and an anti-miR-122 therapy resulted in a significant reduction of circulating cholesterol levels (around 30%). In pigs, miR-122 was identified through sequencing of liver tissue and a lower expression was found in minipigs fed with a high cholesterol diet than those fed with a standard diet, suggesting a potential role of miR-122 in

obesity (Cirera *et al.*, 2010; Wang, Gu and Jiang, 2013). Nowadays, many porcine miRNAs were described to be related with lipid metabolism and/or are involved in meat quality traits (Reddy *et al.*, 2009; Bergen and Burnett, 2013; Li *et al.*, 2017; Huang *et al.*, 2019; Liang *et al.*, 2019; Xing *et al.*, 2019b; Zhang *et al.*, 2019). For example, Xing *et al.* (2019) studied the miRNA liver transcriptome of Landrace pigs with extreme backfat thickness and identified 13 miRNAs differentially expressed between groups. Moreover, Zhang *et al.* (2019) found that changes in miR-21, miR-27a, miR-181a and miR-370 expressions in animals with a diet supplemented with resveratrol are affecting intramuscular fat content.

1.5. Genetic basis of animal breeding

1.5.1. QTLs, GWAS and candidate genes

In the early 1990s, the pig was the first livestock species to which its genome was mapped, with the objective to identify markers linked to quantitative trait *loci* (QTL) (C.S. Haley *et al.*, 1990). A QTL is a position in the genome associated with the variation of a quantitative trait in a population of organisms. On the other hand, the genome wide association study (GWAS) is a study of a genome-wide set of genetic variations in a population and aims to identify the most common genetic variation associated with a specific quantitative trait (Wang *et al.*, 2005). The objective of QTL and GWAS in domestic animals is to identify genes and variants associated with production traits. Molecular markers such as microsatellites or SNPs, which are distributed along the genome, and quantitative phenotypes are used to search QTLs. Usually, molecular markers near or linked to the causal *loci* tends to segregate together because the chance of recombination between them is low. Therefore, the most predictive markers are expected to reside in the proximity of the causal locus. Hence, QTL mapping is a powerful tool to identify genomic regions co-segregating with a specific trait in inter-crossed populations using markers to perform a linkage analysis. The QTL identification depends on the allele diversity that segregates between the parents of the population and the number of recombination events, which requires large information about related individual with known pedigrees (MacKay, Stone and Ayroles, 2009). The

development of GWAS to identify QTLs was carried out since the high-density SNP panels increased the number of genetic markers available. These markers are distributed along the genome and are used to identify a marker allele in linkage or linkage disequilibrium with the causal variant. In addition, due to the use of high-density SNP panels and genetic kinship matrices, instead of using pedigree matrices, GWAS improved the accuracy of QTL analysis, especially if different breeds were studied. Information obtained by GWAS and QTL analyses serves to improve breeding value estimation and to assess genomic selection in the porcine industry.

The first QTL reported in domestic animals was a QTL for fatness on *Sus Scrofa* chromosome 4 (SSC4) in 1994 (Andersson *et al.*, 1994). Since then, several publications reported thousands of QTLs for a broad range of traits in pigs (Figure 1.7.). The Pig QTLdb (Hu *et al.*, 2005) contains 29,865 QTLs/associations from 676 publications, representing 688 different traits (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index; accessed September 2019).

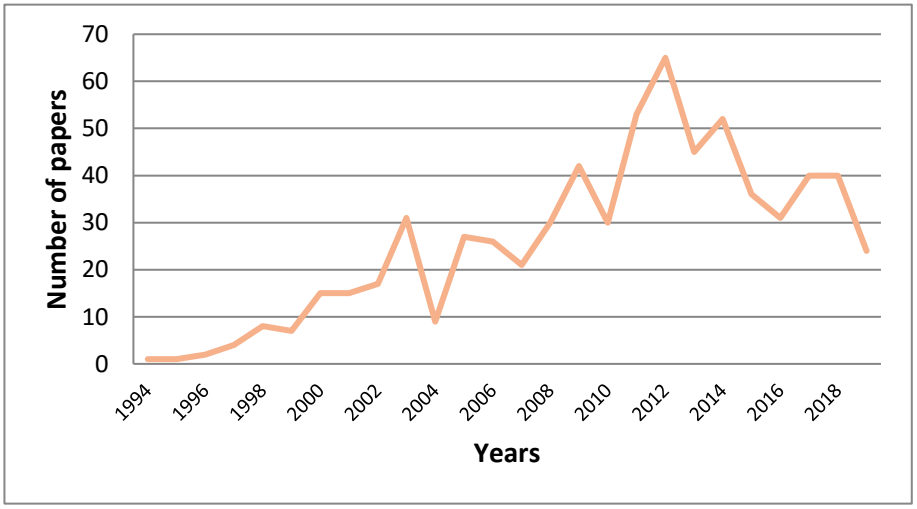


Figure 1.7. Pig QTL publications per year reported from the Pig QTLdb (accessed September 2019).

QTL mapping and GWAS analyses succeeded in the identification of genes containing causal mutations for some QTLs, but their number is still very low because: i) a limited statistical power due to relatively small sample size, ii) production traits are complex

and difficult to measure, iii) genetic variants usually explain a low amount of the genetic variation, and iv) QTL studies conducted in experimental crosses identify large QTL intervals due to the linkage between markers (Goddard and Hayes, 2009). Candidate genes for QTLs have been selected based on both physiological function on the trait and proximity to the QTL for the trait and only some genes have been evaluated for the identification of segregating SNPs and allelic associations with phenotypes.

Some examples of strong candidate genes associated with pig production traits identified in QTL or GWAS analysis were reported and reviewed in several studies (Ernst and Steibel, 2013; Zhang *et al.*, 2014; Muñoz *et al.*, 2018a) and summarized in Table 1.3. Moreover, other candidate genes related to androstenone concentration, eating behaviour, farrowing and haematological traits in pigs were reviewed in Sharma *et al.* 2015 (Sharma *et al.*, 2015). Within the candidate genes, transcription and control regulators have a prominent interest due to their role in gene regulation.

Table 1.3. Summary of candidate genes identified in QTLs or GWAS analysis for pig production traits.

Trait	Candidate Genes
Coat colour	*KIT, MC1R, TYRP1*
Growth, fatness and carcass composition	*ADIPOQ, FASN, FTO, IGF2, LEP, LEPR, MC4R, MRF, MSTN, MYPN, POU1F1, PLAG1, TAS2R39, TAS2R4*
Meat quality	*ACACA, ACSL4, CAPNS1, CAST, CA3, CYBSA, CYP2E1, ELOVL6, FABP4, FABP5, PCK1, PHKG1 PPARGC1A, PRKAG3, RYR1, SCD, TTN*
Litter size	*AHR, ESR1, FSHB, PRLR, RBP4*
Disease resistance	*FUT1, GBP5, MUC4, NRAMP, SLA*

1.5.2. eQTL mapping

An expression QTL (eQTL) is a genomic position associated with gene expression differences. The QTL co-localization with an eQTLs is a powerful tool to identify candidate genes to explain a particular trait. Moreover, eQTL mapping can reveal gene regulatory networks and key regulators for the phenotypic variance (Verdugo *et al.*, 2010; Ernst and Steibel, 2013).

The eQTL analysis allows to discriminate between *cis* and *trans*-acting mode of action and led to identify hotspot *loci* and regulators. The *cis*-eQTL is a genetic variant mapped close or inside the studied gene and directly affects its gene expression level. In contrast, a *trans*-acting eQTL is a genetic variant mapped in a different genomic location of the studied gene, and may indirectly affect the target gene expression (Cheung and Spielman, 2009) (Figure 1.8.).

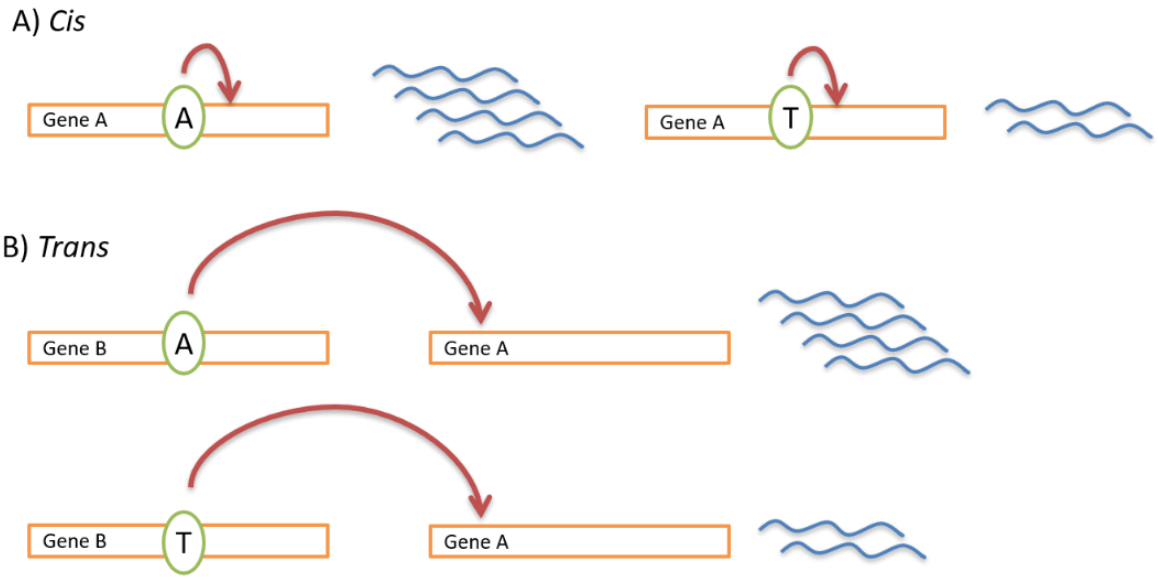


Figure 1.8. Representation of A) *cis*- and B) *trans*-acting eQTLs regions. In the *cis*-eQTL the expression of a gene located close to the SNP varies according to the presence of one allele, whereas for a *trans*-eQTL the SNP influencing the target gene expression level is located far away.

Cis-acting eQTLs usually explain a large fraction of the variance in the gene expression and have a high interest. *Trans*-eQTLs are often regulating a large number of genes, and are reported as regulatory hotspots (Schadt *et al.*, 2003).

The first eQTL mapping studies were done in humans, plants and model organisms and were published during the early 2000s (Jansen and Nap, 2001; Schadt *et al.*, 2003). However, so far there have been few eQTL studies in livestock animals and only a few in pigs due to the high cost and complexity of performing eQTL analysis. To date, the most common studies are using transcriptomic data of skeletal muscle and have reported eQTL studies for production traits in pigs (Table 1.4.). Moreover, published studies of our group analysed the muscle (Puig-Oliveras *et al.*, 2016), liver (Ballester *et al.*, 2017b) and adipose tissue (Revilla *et al.*, 2018) of a subset of lipid-related genes. Using eQTL analysis some candidate genes and genetic networks were described.

Table 1.4. Summary of eQTL studies for genes associated with growth, fatness and meat quality production traits in pigs.

Related trait	References
Growth	(Steibel *et al.*, 2011; Ponsuksili *et al.*, 2012)
Fatness and fatty acid composition	(Ponsuksili *et al.*, 2011; Steibel *et al.*, 2011; Cánovas *et al.*, 2012; Muñoz *et al.*, 2013a; Martínez-Montes *et al.*, 2017; Revilla *et al.*, 2018; González-Prendes *et al.*, 2019)
Meat quality	(Ponsuksili *et al.*, 2008, 2010, 2014; Wimmers, Murani and Ponsuksili, 2010; Steibel *et al.*, 2011; Heidt *et al.*, 2013; Muñoz *et al.*, 2013a; Pena *et al.*, 2013; Manunza *et al.*, 2014; Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017b; González-Prendes *et al.*, 2017; Velez-Irizarry *et al.*, 2019)

1.6. The IBMAP cross

The IBMAP consortium was created in 1996 with the collaboration among the *Universitat Autònoma de Barcelona* (UAB), the *Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria* (INIA), and the *Institut de Recerca i Tecnologia Agroalimentàries* (IRTA). Different backcrosses were made between Iberian pig's boars, which have an excellent meat quality, and sows of other three breeds: Landrace, Duroc and Pietrain. The F1 obtained from the three different crosses were backcrossed again with sows of their respective maternal line. Moreover, other F2 and F3 crosses were performed (Figure 1.9.).

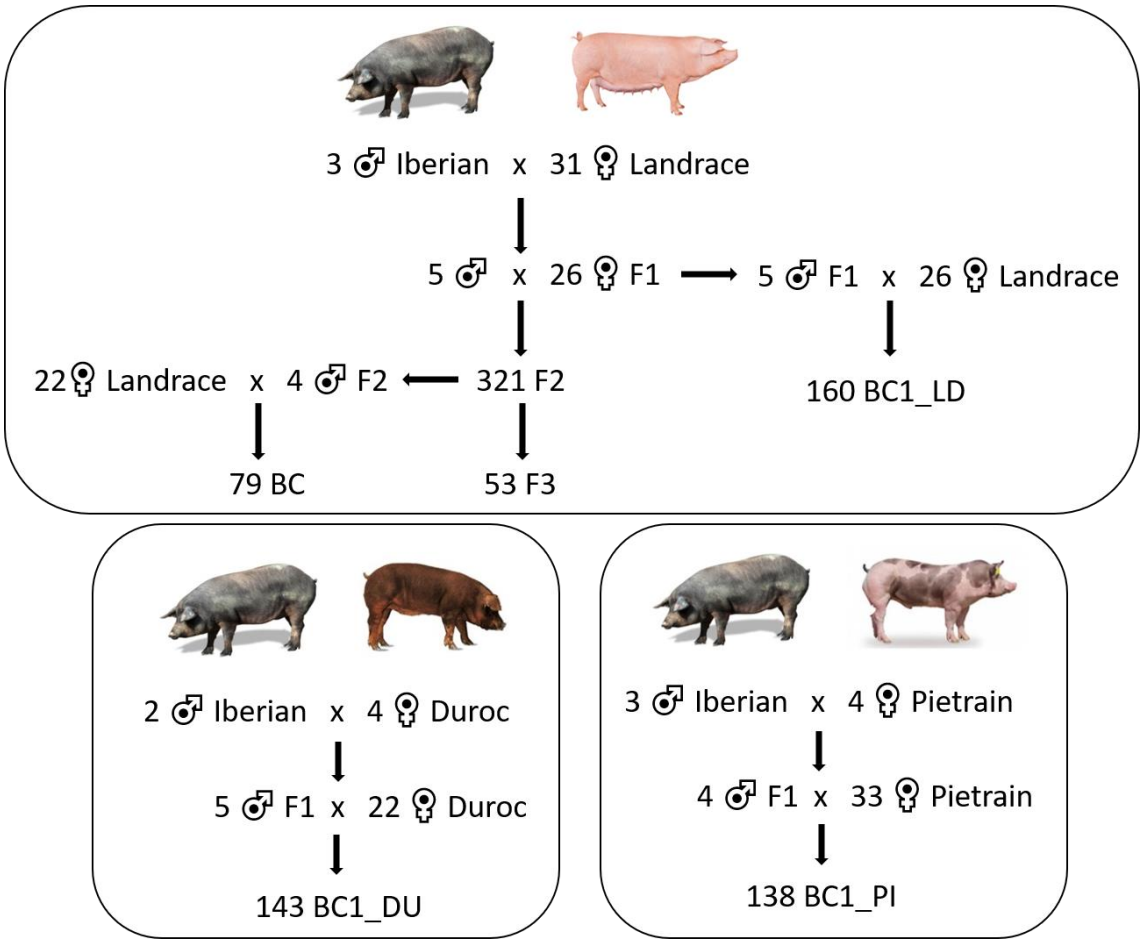


Figure 1.9.: Schematic representation of the three IBMAP backcrosses (BC1_LD, BC1_DU and BC1_PI).

Different pig breeds were chosen because they differ in meat quality, growth, fatness, prolificacy and feed efficiency traits. The Iberian pig is a local and rustic breed

produced in Spain and it is well known for its excellent meat quality and cured products, with a higher content of SFA and MUFA (mainly oleic acid), but with a lower productive conformation than commercial breeds (Serra *et al.*, 1998).

On the contrary, Landrace is a lean international breed with less intramuscular fat and higher content of PUFA. It suffered a strong selection for production traits presenting high prolificacy and growth. In addition, Duroc breed is characterized by its rusticity, with good conformation and low food consumption. Although they have low growth rates and tend to be fat, they are used to improve carcass meat quality. Finally, Pietrain pigs present an excellent carcass conformation conferring a better production of lean meat and an efficient conversion rate. However, they are less prolific than other commercial breeds such as Landrace (Kouba and Sellier, 2011).

1.6.1. Identification of candidate genes of QTLs in the IBMAP cross

The main objective of the IBMAP consortium was to identify QTLs associated with pork meat quality and growth traits. The first studies, performed on the Iberian x Landrace F2-cross, were based on microsatellites markers and identified significant associated regions for carcass quality, growth, fatness and fatty acid composition measured in backfat, on chromosomes SSC2, SSC3, SSC4, SSC6, SSC7, SSC8, SSC10, SSC12, and SSCX (Óvilo *et al.*, 2000, 2002; Pérez-Enciso *et al.*, 2000, 2002, 2005; Clop *et al.*, 2003; Mercadé *et al.*, 2005, 2006; Muñoz *et al.*, 2007).

Afterwards, new technologies such as the PorcineSNP60 BeadChip of *Illumina*, were used to improve the resolution of the previous QTLs described and to find new genomic regions associated with the analysed traits (Fernández *et al.*, 2012; Y. Ramayo-Caldas *et al.*, 2012; Corominas *et al.*, 2013; Muñoz *et al.*, 2013; Revilla *et al.*, 2014). Within these QTLs some functional candidate genes were identified for growth, fatty acid composition, fatness and intramuscular fat content traits in the IBMAP population: *ACACA* (Muñoz *et al.*, 2007), *ACADM* (Kim *et al.*, 2006), *ACSL4* (Corominas *et al.*, 2012), *APOA2* (Ballester *et al.*, 2016), *CDS1* and *CDS2* (Mercadé, Sánchez and Folch, 2007), *DECR* (Clop *et al.*, 2002), *DGAT1* (Mercadé, Sánchez and Folch, 2005), *ELOVL6* (Corominas *et al.*, 2013b, 2015), *FABP2* (Estellé *et al.*, 2009b), *FABP3* (Óvilo *et al.*, 2002), *FABP4* (Mercadé *et al.*, 2006), *FABP5* (Estellé *et al.*, 2006), *FASN* (Muñoz *et*

al., 2007), *GIP* (Muñoz *et al.*, 2007), *IGF2* (Estellé *et al.*, 2005; Criado-Mesas *et al.*, 2019), *LEPR* (Óvilo *et al.*, 2005; Muñoz *et al.*, 2009), *MTTP* (Estellé *et al.*, 2009a), *MAML3* and *SETD7* (Revilla *et al.*, 2014).

Furthermore, the RNA-Seq technique was used to identify differential expressed genes in the three main metabolic tissues: liver (Ramayo-Caldas *et al.*, 2012a), adipose tissue (Corominas *et al.*, 2013) and muscle (Puig-Oliveras *et al.*, 2014) in BC1_LD animals. Two extreme groups for fatty acid composition were selected for transcriptomic analysis. In the three RNA-Seq studies common pathways were found related with LXR/RXR, peroxisome proliferator activated receptors (PPARs) and fatty acid β-oxidation pathways. In particular, *SCD* gene was differentially expressed in backfat and muscle, while *ELOVL6* and *FASN* were differentially expressed in backfat.

Finally, GWAS and haplotype association analyses using the three different genetic backgrounds, BC1_LD, BC1_DU and BC1_PI, identified nine QTL regions for growth, premier cut weights and intramuscular fat, as well as, some backcross specific QTL regions. In this study, six strong candidate genes were identified (Martínez-Montes *et al.*, 2018).

More recently, the expression of several candidate genes was analysed in muscle (Puig-Oliveras *et al.*, 2016), liver (Ballester *et al.*, 2017b), and backfat (Revilla *et al.*, 2018) tissues by RT-qPCR in the BC1_LD animals. In muscle tissue, the *NR3C1* transcription factor was proposed to be a major regulator in fatty acid metabolism, and this and other genes were found co-localizing with QTLs for fatness and growth traits (*ARHGAP6*, *IGF2*, *MC2R*, and *MGLL*) (Puig-Oliveras *et al.*, 2016). In liver, the *NR3C1* gene was also identified as a potential regulator. In addition, a hotspot on SSC8 was associated with the expression of eight genes and *TBCK* gene was proposed as a master regulator (Ballester *et al.*, 2017b). In backfat adipose tissue three *cis*-eQTLs were found for *ACSM5*, *FADS2* and *FABP4*, where *SREBF1* and *PPAR* were described as gene expression regulators, while a significant *trans*-eQTL for *ELOVL6* was also associated with the expression of *ELOVL5* and *SCD* genes (Revilla *et al.*, 2018).

Objectives

Chapter 2

This PhD thesis was done under the framework of the AGL2014-56369-C2-2-R and AGL2017-82641-R projects funded by the *Ministerio de Economía y Competitividad* (MINECO). The animal material used in the current research was generated by the IBMAP consortium involving INIA, IRTA and UAB research groups.

The main objective was to study the genetic and molecular basis determining fatty acid composition in pigs.

The specific objectives of this thesis were:

1. To deepen into the study of the expression and regulation of 45 lipid-related genes in the *Longissimus dorsi* muscle of pigs from three different genetic backgrounds to evaluate differences in gene expression and its regulation within and across populations.

2. To analyse the effect of *IGF2:g.3072G>A* polymorphism on adipose tissue *IGF2* gene expression, its regulation and fatty acid composition.

3. To study the expression of the porcine miR-33a and miR-33b in liver, muscle and adipose tissue and their association with fatty acid composition, to better understand the regulation and role in lipid metabolism of these miRNAs.

4. To characterize the transcriptome architecture of the porcine *Longissimus dorsi* muscle by RNA-Seq and to identify potential regulators of muscle gene expression.

Papers and Studies

Chapter 3

PAPER I

Identification of eQTLs associated with lipid metabolism in *Longissimus dorsi* muscle of pigs with different genetic backgrounds

Lourdes Criado-Mesas[1]*, Maria Ballester[2], Daniel Crespo-Piazuelo[1,3], Anna Castelló[1,3], Ana I. Fernández[4] and Josep M. Folch[1,3]

[1]Departament de Genòmica Animal, Centre de Recerca en Agrigenòmica (CRAG), CSIC-IRTA-UAB-UB, Barcelona, Spain.

[2]Departament de Genètica i Millora Animal, Institut de Recerca y Tecnologia Agraroalimentàries (IRTA), Caldes de Montbui, Spain.

[3]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, UAB, Bellaterra, Spain.

[4]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain.

* Corresponding author

Scientific Reports (in revision)

Abstract

Intramuscular fat content and its fatty acid composition affect porcine meat quality and its nutritional value. The present work aimed to study and validate the genetic basis of the expression of 45 genes involved in lipid metabolism in the porcine muscle *(Longissimus dorsi)* of three different experimental backcrosses based on the Iberian breed. Expression genome-wide association studies (eGWAS) were performed between the muscle gene expression values, measured by real-time quantitative PCR, and the genotypes of 38,426 SNPs distributed along all chromosomes. The eGWAS identified 186 eSNPs located in ten *Sus scrofa* regions and associated with the expression of *ACSM5*, *ACSS2*, *ATF3*, *DGAT2*, *FOS* and *IGF2* (FDR<0.05) genes. Two expression quantitative trait loci (eQTLs) for *IGF2* and *ACSM5* were classified as *cis*-acting eQTLs, suggesting a mutation in the same gene affecting its expression. Conversely, ten eQTLs showed *trans*-regulatory effects on gene expression. When the eGWAS was performed for each backcross independently, only three common *trans*-eQTL regions were observed, indicating different regulatory mechanisms or allelic frequencies among the breeds. In addition, hotspot regions regulating the expression of several genes were detected. Our results provide new data to better understand the functional regulatory mechanisms of lipid metabolism genes in muscle.

Introduction

Studies on the traits that determine the quality of pork meat and their derived products have received increasing attention in recent years. The intramuscular fat (IMF) content and its fatty acid (FA) composition are considered determinant for meat quality, playing a central role in the nutritional values of the meat[1]. IMF influences meat flavour, juiciness, tenderness and firmness, which are important traits for consumer acceptance. On the other hand, its FA composition will determine how healthy is the product since it is well-known that some FAs are essential for humans, such as ω-3 and ω-6 polyunsaturated FAs (PUFAs)[2].

During the last years, pig breeding companies have produced commercial pigs that grow faster and have superior carcasses. However, these carcasses have become leaner having less IMF and, therefore, producing a decrease in the meat quality according to consumers. Otherwise, local breeds such as the Iberian pig present a high-fat deposition and FA desaturation values and have a special interest in the production of high-quality dry-cured cuts, such as loin and ham[3]. Often the Iberian pig is crossed with other breeds to improve its reproductive and growth traits, although crossing has been associated with a decrease in meat quality[4].

Several studies agree that genetic factors can determine intramuscular FA composition in pigs[1,5–7]. For example, significant breed effects have been reported for IMF, water binding capacity, colour and tenderness. Thus, differences according to the genetic background have made the industry aware of it when improving the meat quality of pork[8].

In recent years, genome-wide association studies (GWAS) have been used to detect genetic variants involved in FA composition traits, unravelling the complex genetic basis of these quantitative traits[9–14]. In general, genes involved in pathways or functions related to lipid metabolism are regulated at the transcriptional level, and studies conducted on the molecular mechanisms controlling these functions help to understand the genetic basis of traits related to FA composition in muscle tissue[15]. In previous studies, we have identified differentially expressed (DE) genes in the muscle transcriptome among two groups of extreme animals for FA composition in an Iberian

x Landrace cross by RNA-Seq, reinforcing the view that variation in gene expression and its genetic basis may play an important role in the genetic determinism of these traits[16]. In addition, an expression genome-wide association study (eGWAS) of 45 lipid-related genes in the muscle of 114 Iberian × Landrace animals allowed the identification of genomic regions regulating the expression of these genes[17]. Two other gene expression studies related to lipid metabolism were performed in liver and backfat in the same experimental population[18,19].

The main goal of the present work was to study and validate the expression and regulation of a selected set of 45 genes involved in lipid metabolism in the porcine *Longissimus dorsi* (LD) muscle in a total of 355 animals belonging to three different backgrounds. Specifically, the eGWAS study was conducted for: a) data generated in Puig-Oliveras *et al*. (2016)[17] from the BC1_LD (25% Iberian and 75% Landrace) population, and re-analysed in the present study using the *Sscrofa* 11.1 genome assembly, and b) data generated in the current study from the BC1_DU (25% Iberian and 75% Duroc) and BC1_PI (25% Iberian and 75% Pietrain) populations, to evaluate differences in gene expression and its regulation within and across populations.

Results and discussion

Sex and genetic background effect on gene expression

A sex bias in the expression of genes associated with lipid metabolism has been previously described in muscle and other tissues such as liver[20,21]. Hence it is relevant to understand the mechanisms of sex-differential gene expression.

In the global study, including the three backcrosses (3BCs), 30 out of the 45 genes presented significant sex effect (*p*-value ≤ 0.05) on gene expression: *ACSM5*, *ACSS1*, *ACSS2*, *ANGPT1*, *AQP7*, *ATF3*, *CREG1*, *CROT*, *DGAT2*, *ETS1*, *HIF1AN*, *IGF2*, *LXRA*, *NCOA1*, *NCOA2*, *NCOA6*, *NFKB*, *PIK3R1*, *PLA2G12A*, *PPARA*, *PPARD*, *PPARG*, *PPARGC1A*, *PRKAA1*, *PXMP3*, *RXRG*, *SCD*, *SETD7*, *SP1* and *SREBP1C* (Figure 1). In general, there were more genes over-expressed in females, 24 out of 30, than in males. Six genes presented higher expression in males: *ACSS1*, *ATF3*, *ETS1*, *PPARA*, *PPARD* and *PPARGC1A*, being some of them relevant regulators implicated in lipolytic pathways.

Genes over-expressed in females were implicated in transcriptional regulation and control (*CREG1*, *LXRA*, *NFKB1*, *NCOA1*, *NCOA2*, *NCOA6, PPARG, PRKAA1*, *RXRG*, *SP1* and *SREBP1c*), FA β-oxidation (*CROT*, *PXMP3* and *SCD*), lipid storage (*ACSM5*, *DGAT2*, *HIF1AN* and *AQP7*), cholesterol (*ACSS2*, *ANGPT1* and *SETD7)* and the AKT pathway (*IGF2*, *PIK3R1* and *PLA2G12A*). In addition, the *IGF2* gene, which has been involved in muscle growth and fat deposition[22], showed a higher expression in females.

Overall these results are in accordance with previous studies describing differences in fat distribution and lipid metabolism between males and females. In humans, males tend to present higher activity in lipolytic pathways, while females present higher rates of lipogenesis and accumulation of triglycerides, so they have a higher risk to gain fat and develop obesity[23]. In a similar way, female pigs seem to develop obesity more readily than male pigs[24].

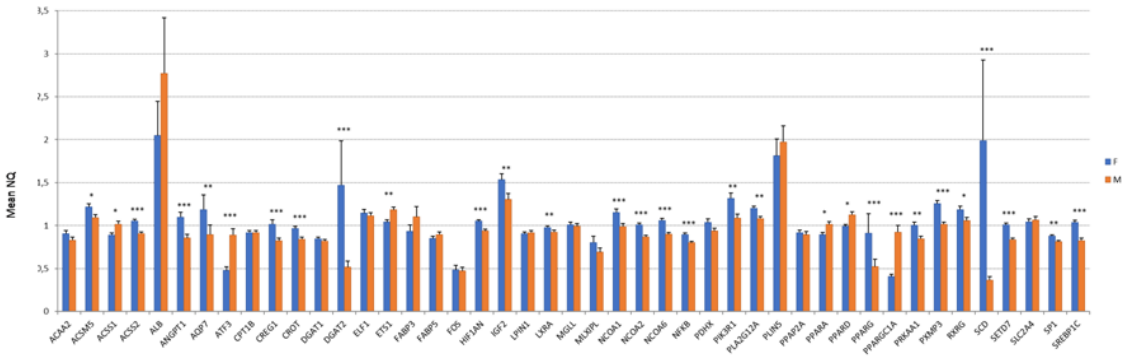


Figure 1: Comparison between females (F) and males (M) of mRNA expression levels of 45 lipid-related genes in animals from the 3BCs. Data are presented as mean ± standard error of the mean (SEM). Significant differences are labelled as * *p*-value ≤ 0.05, ** *p*-value ≤ 0.01 and *** *p*-value ≤ 0.001.

Among the list of sex-biased genes, it is worth to highlight the role of *SREBP1C* and *PPARA* as key regulatory genes for lipid metabolism. Their differential sex expression pattern was also observed in liver and adipose tissue of BC1_LD animals. *PPARA* gene showed a higher expression in the liver and adipose tissue of females, in contrast to muscle. The expression of *SREBP1C* in adipose tissue and liver, as occurred in muscle, was higher in females[17–19]. The *SREBP1C* is a transcription factor that regulates the

expression of a broad range of lipid metabolism genes[25] which agrees with the higher number of genes over-expressed in females.

A breed effect on the expression of genes involved in energy balance and lipogenesis was reported in a comparison between Iberian and Duroc pigs[26]. In our study, a significant backcross effect (*p*-value≤ 0.05) on gene-expression levels was detected in 37 out of the 45 genes analysed: *ACAA2*, *ACSS1*, *ACSS2*, *ALB*, *ANGPT1*, *AQP7*, *MLXIPL*, *CPT1B*, *CREG1*, *CROT*, *DGAT1*, *DGAT2*, *ELF1*, *ETS1*, *FABP5*, *FOS*, *HIF1AN*, *IGF2*, *LXRA*, *MGLL*, *NCOA2*, *NFKB*, *PDHX*, *PIK3R1*, *PLIN5*, *PPARA*, *PPARD*, *PPARG*, *PPARGC1A*, *PRKAA1*, *PXMP3*, *RXRG*, *SCD*, *SETD7*, *SLC2A4*, *SP1* and *SREBP1C* (Figure 2). Overall, 18 and 16 out of 45 genes were over-expressed in BC1_LD and BC1_DU respectively and are involved in a wide range of functions. In summary, genes more related to lipogenic pathways were more expressed in BC1_LD whereas genes related to lipolytic pathways were higher expressed in BC1_DU. Finally, 3 out of 45 genes were over-expressed in BC1_PI and were mainly related to transcriptional regulation and control.

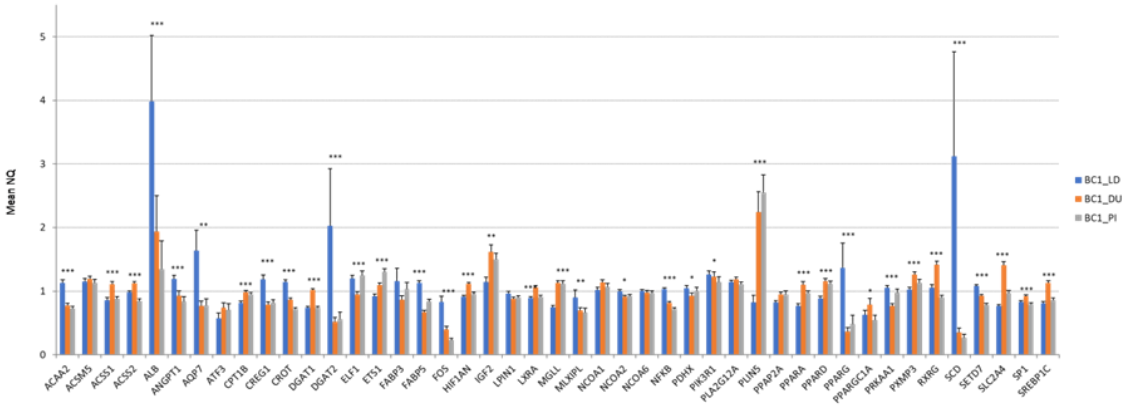


Figure 2: Comparison between the three experimental backcrosses in the mRNA expression levels of 45 lipid-related genes. Data represents means ± standard error of the mean (SEM). Significant differences are labelled as * *p*-value ≤ 0.05, ** *p*-value ≤ 0.01 and *** *p*-value ≤ 0.001.

Altogether these results indicated an effect of sex and breed on gene expression, therefore they were considered in association studies and included as co-factors in the model.

Gene expression correlations

In order to identify co-expression patterns in the selected genes analysed in our study, a co-expression network using PCIT algorithm[27] was performed with the expression data of 3BCs animals (Figure 3).

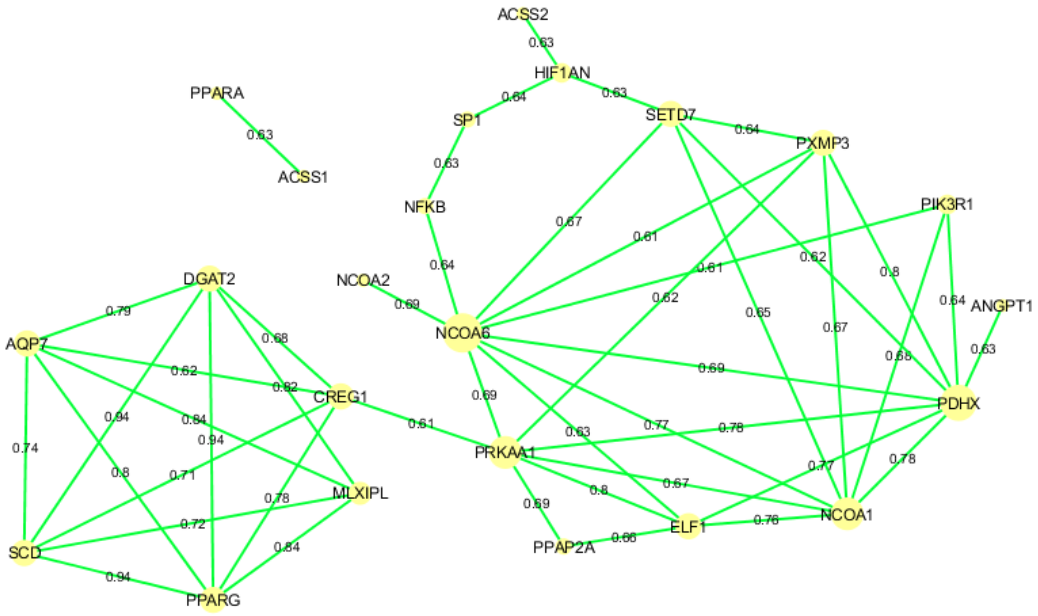


Figure 3: Gene co-expression network in 3BCs using the PCIT algorithm[27]. After filtering by significance and r≥0.6, 23 from the 45 initial genes are shown. Node size represents the degree of a node.

Two groups of co-expressed genes were identified by PCIT algorithm. It is particularly interesting to mention the strongest correlations found for *SCD*, *PPARG*, and *DGAT2* genes in the first group and which were previously identified in the BC1_LD study[17]. *CREG1* and *PRKAA1* were identified linking both groups of co-expressed genes. Remarkably, among this second group of co-expressed genes strong correlations for

ELF1, *NCOA1*, *NCOA6*, *PDHX*, *PRKAA1*, *PXMP3* and *SETD7* were identified and the highest node degree corresponded to *NCOA6* and *PDHX*.

Genome-wide association studies for gene expression and eQTL identification

An eGWAS was performed with the muscle gene expression values and the genotypes of 38,426 single nucleotide polymorphisms (SNPs) distributed along all chromosomes in 355 3BCs animals. The eGWAS identified 186 expression-SNPs (eSNPs) located in 10 *Sus scrofa* chromosomes (SSC) regions of SSC1, SSC2, SSC3, SSC6, SSC7, SSC11, SSC13 and SSC16 and associated with the expression of *ACSM5*, *ACSS2*, *ATF3*, *DGAT2*, *FOS* and *IGF2* (FDR<0.05) genes (Supplementary Table S1). Ten eQTLs showed *trans*-regulatory effects on gene expression and two of them, *IGF2* and *ACSM5*, were also classified as *cis*-acting, suggesting that there is a mutation in the same gene or in a proximal genomic region affecting its expression (Table 1). Both *cis* and *trans*-eQTLs were represented in Figure 4.

Inter-val	Gene	Chr.	Start Position (bp)	End Position (bp)	Size (Mb)	SNPs	Type of eQTL	Candidate genes
1	*ACSM5*	3	18,557,492	53,699,303	35.14	58	*cis/ trans*	*ACSM5* and *IL4R*
2	*ACSS2*	6	17,315,441	17,502,570	0.19	2	*trans*	
3	*ACSS2*	7	111,283,606	112,227,872	0.94	8	*trans*	
4	*ACSS2*	13	156,576,634	156,644,710	0.07	2	*trans*	
5	*ATF3*	1	181,624,438	181,702,614	0.08	3	*trans*	
6	*ATF3*	13	177,313,258	177,546,824	0.23	2	*trans*	
7	*DGAT2*	16	2,764,727	2,779,416	0.01	2	*trans*	
8	*FOS*	1	0	493,510	0.49	2	*trans*	
9	*FOS*	11	8,855,571	19,677,423	10.82	3	*trans*	*RB1* and *FOXO1*
10	*IGF2*	2	1,000,000	25,964,207	24.96	104	*cis/ trans*	*IGF2, SF1* and *NR1H3*

Table 1: Significant eQTLs for the 45-muscle gene expression study in 3BCs animals. Start and end positions refer to the eQTL interval and are based on *Sscrofa* 11.1 assembly. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval. For the *cis*-eQTLs regions only the analyzed gene was annotated as positional candidate gene.

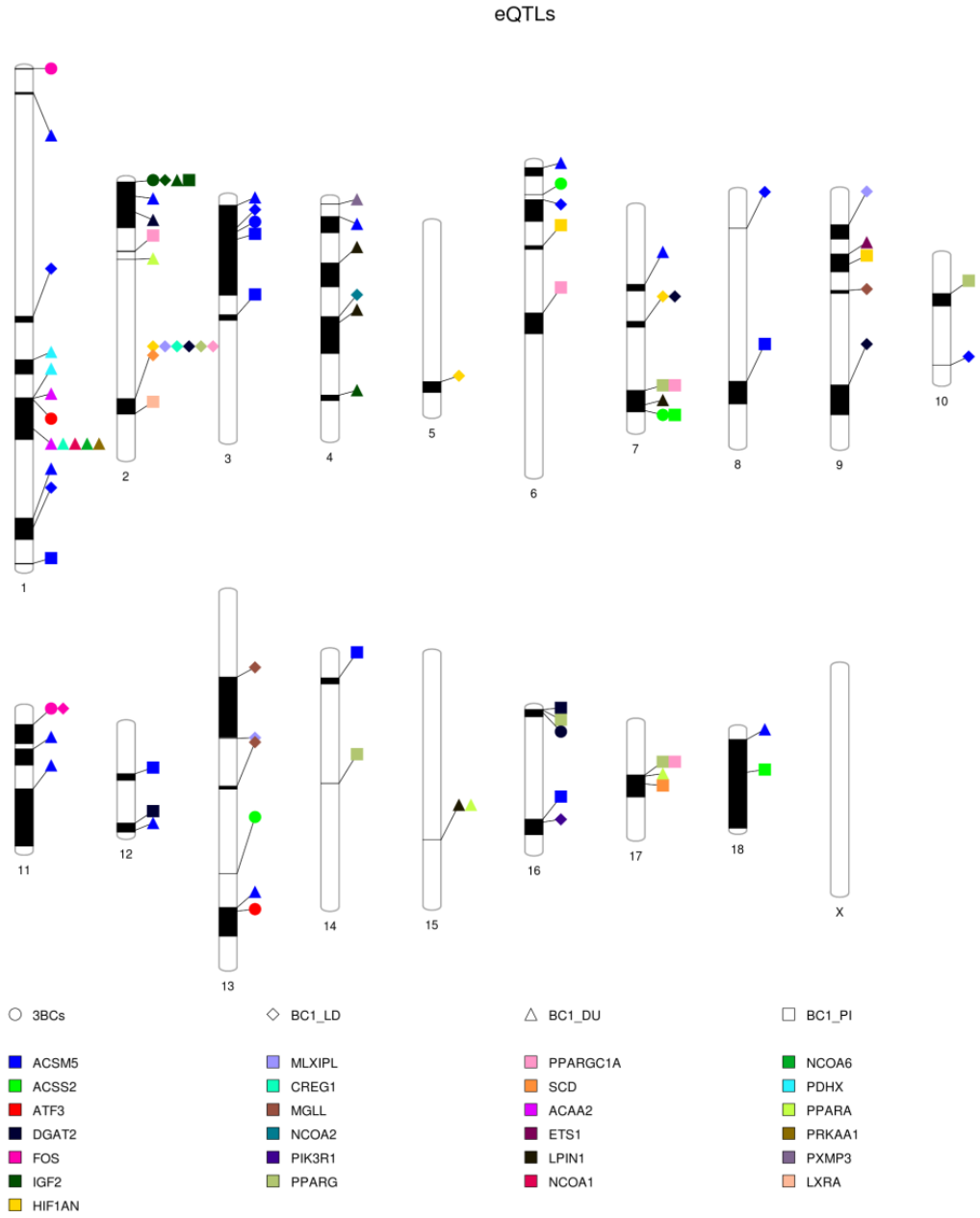


Figure 4: PhenoGram plot representing associated gene expression regions along pig chromosomes in the 3BCs study and in each backcross individually. The shape indicates the backcross or the 3BCs altogether and the colour indicates the gene name as it is indicated in the legend.

Cis-eQTLs:

For the *IGF2* *cis*-eQTL region, the *IGF2g.3072G>A* SNP was the most significantly associated polymorphism (*p*-value=3.24x10^{-44}) and explained the 70% of the muscle *IGF2* expression variance, approximately (Figure 5).

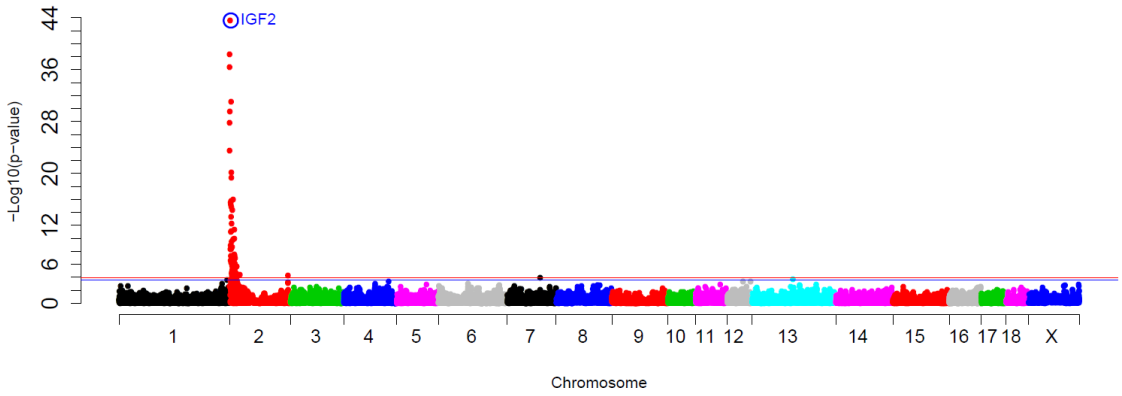


Figure 5: GWAS plot of muscle *IGF2* gene expression in the 3BCs study. Chromosome positions in Mb based on *Sscrofa* 11.1 assembly of the pig genome are represented in the X-axis and the –log10 (*p*-value) is on the Y-axis. Horizontal lines represent the genome-wide significance level (FDR-based *q*-value < 0.1 corresponds to blue line and FDR-based *q*-value < 0.05 to red line). The *IGF2*:g.3072G>A polymorphism is circled and labelled as IGF2 in colour blue.

The *IGF2:g.3072G>A* substitution has been identified as the causal mutation of an imprinted QTL for muscle growth, fat deposition and heart size[22] and it is maternally imprinted in most animal tissues[27]. The *IGF2g.3072G>A* mutation is located in a well-conserved CpG island, which is hypomethylated and abrogates the binding site for an *IGF2* transcriptional repressor called ZBDE6, leading to a three-fold up-regulation of the *IGF2* expression in pig skeletal muscle[22,28].

An imprinting model was tested for muscle gene expression in 327 animals in which the paternal allele was deduced from progenitor's genotypes (Figure 6). Animals with the paternally-inherited *A* allele (A^P) of the *IGF2:g.3072G>A* polymorphism showed the highest *IGF2* gene expression in muscle (*AA*: NQ mean=2.29, n=130 and A^PG^M: NQ mean=2.65, n=26) compared to animals with paternally-inherited G allele (A^MG^P: NQ mean=0.65, n=122 and *GG*: NQ mean=0.78, n=76).

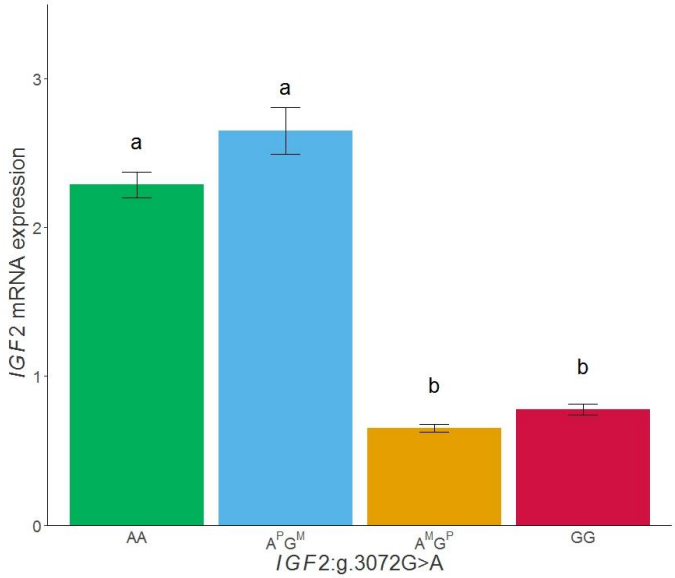


Figure 6: Plot of mRNA expression values (NQ) of *IGF2* in muscle tissue according to the *IGF2*:g.3072G>A genotype. A^PG^M indicates a paternally inherited A allele and maternal inherited *G* allele, on the contrary, A^MG^P represents a maternal inherited A allele and paternal inherited *G* allele. Data represents means ± standard error of mean (SEM). Values with different superscript letters (a, b) indicate significant differences between groups (*p*-value < 0.05).

Therefore, the *IGF2*:g.3072G>A SNP genotype and the imprinting model explained the differences observed in *IGF2* gene expression in muscle, being the *IGF2* genetic variant the major regulator of gene expression in muscle in different genetic backgrounds (see below specific data for each backcross).

A previous study of our group reported that *IGF2* polymorphism was also the most significant associated SNP with *IGF2* mRNA expression in adipose tissue[19], but it explained only 25% of the phenotypic variance compared to the 70% explained in muscle tissue, suggesting that other genetic variants, potentially *trans*-regulation as reported in the current study, may affect the gene expression in adipose tissue. Nevertheless, the *IGF2* gene expression followed a maternal imprinting model in both tissues[29].

The *ACSM5* gene, target of the other *cis*-eQTL region identified, is involved in pathways such as conjugation of carboxylic acids and FA beta-oxidation. A SSC3 *cis*-eQTL was reported in a previous study of our group analysing the *ACSM5* expression in BC1_LD

population[17]. The *ACSM5* proximal promoter region was amplified and sequenced in ten BC1_LD animals and subsequently three polymorphisms were found. The most proximal 5' mutation, *rs331702081* (hereinafter known as *ACSM5.P*) was the most significantly associated SNP with the *ACSM5* gene expression in the BC1_LD population[17]. Thus, in the current study the *ACSM5.P* was genotyped in the BC1_DU and BC1_PI populations.

In the eGWAS with all three backcrosses the *ACSM5.P* SNP presented the strongest association with muscle *ACSM5* gene expression (*p*-value=1.39x10^{-27}) (Figure 7). The polymorphism located in the promoter region explained approximately the 40% of the phenotypic variance, suggesting the presence of additional genetic factors regulating its gene expression (see below specific data for each backcross). Further analysis should be done to understand the transcriptional regulation of *ACSM5* gene.

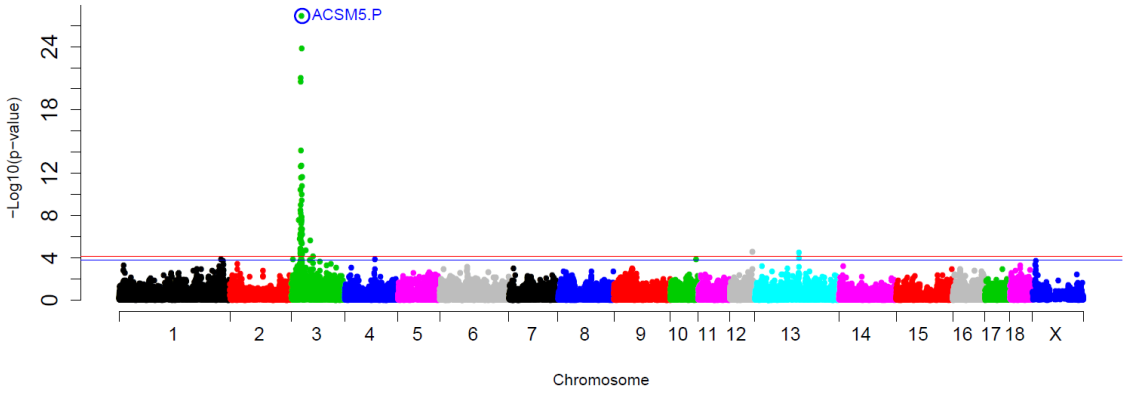


Figure 7: GWAS plot of muscle *ACSM5* gene expression in the 3BCs study. Chromosome positions in Mb based on *Sscrofa* 11.1 assembly of the pig genome are represented in the X-axis and the –log10 (*p*-value) is on the Y-axis. Horizontal lines represent the genome-wide significance level (FDR-based *q*-value < 0.1 corresponds to blue line and FDR-based *q*-value < 0.05 to red line).

In a previous study of our group the *ACSM5.P* mutation has been also described as the most significantly associated SNP with *ACSM5* gene expression in backfat adipose tissue of the BC1_LD population[19]. Nonetheless, the correlation between the *ACSM5* gene expression in backfat and muscle was 0.60, suggesting that the gene expression in both tissues could be regulated by different genetic variants. In addition, two

transcription factors (*ARNT* and *STAT6*) that bind only when the *A* allele is present were identified[19]. Hence, genetic variation on the promoter region of *ACSM5* could be a key regulator of the *ACSM5* gene expression, at least in muscle and adipose tissue.

Trans-eQTLs:

A total of 783 genes were located in the 10 *trans*-eQTL genomic regions identified in our study. Among them, we identified potential lipid metabolism regulatory genes in three regions (Table 1: interval 1, 9 and 10). The *ACSM5* eGWAS revealed a *trans*-eQTL located in the 18.5 Mb – 53.6 Mb region of SSC3, where the Interleukin 4 Receptor (*ILR4*) gene was mapped. Polymorphisms in *ILR4* have been associated with high density lipoprotein-cholesterol levels, suggesting the possible role of *IL4R* gene in lipid metabolism in humans[30]. The FBJ Murine Osteosarcoma Viral Oncogene Homolog (*FOS*) eGWAS revealed a *trans*-eQTL in the 8.9 Mb - 19.7 Mb region of SSC11, where a gene involved in lipid metabolism was mapped: Forkhead Box O1 (*FOXO1*). From the *FOXO* TF family, *FOXO1* is the isoform with the highest expression in muscle and has been proposed as a regulator of energy metabolism and the insulin signalling pathway[31]. It is also involved in muscle differentiation and can interact with other transcription factors such as *PPARG* and *HNF4A* to regulate insulin gene expression and IMF accumulation[32]. Moreover, *FOXO1* was found to regulate *FOS* gene expression in skeletal muscle, increasing their levels during cancer cachexia in humans[33]. Retinoblastoma 1 (*RB1)* gene was also a transcription factor mapped in this region and is involved in gene expression control. *RB1* plays an important role in cell cycle and cell differentiation and is also considered as a key regulator during adipogenesis. However, it is highly expressed in muscle tissue probably due to its role in muscle differentiation[34]. In humans, *RB1* was found co-expressed with *FOS* gene and is involved in proliferation and apoptosis in myosarcoma[35]. A prediction of a functional integration network was done by GeneMANIA, showing a gene co-expression between *FOS* and *FOXO1*, a predicted functional gene relationship between *FOS* and *RB1*, and *FOXO1* with *PPARG* and *HNF4*, protein-protein interactions among *FOXO1* and *RB1* and finally a *FOS*, *PPARG* and *RB1* gene pathway.

The Splicing Factor 1 (*SF1*) gene was mapped in the *IGF2 trans*-eQTL region located on SSC2 (Table 1) and it was previously described as a candidate gene for *IGF2* regulation

in adipose tissue[29]. A member of the *LXR* nuclear receptor family named nuclear receptor subfamily 1 group H member 3 (*NR1H3*) was also mapped in this *trans*-eQTL region and chosen as a possible candidate gene due to its involvement in the deposition of lipids in pigs, which may affect lean muscle fat content[36].

The rest of the *trans*-eQTL regions were identified for *ACSS3* (SSC6, SSC7 and SSC13), *ATF3* (SSC1 and SSC13), *DGAT2* (SSC16) and *FOS* (SSC1). However, no candidate regulator genes could be identified in these genomic regions. This may be explained by the small intervals size, the lack of gene information in the pig assembly or the presence of other regulators such as enhancers, miRNAs and long-non-coding RNAs among others.

eGWAS analysis for each backcross independently

Expression-GWAS studies were also performed for each backcross independently and 420, 420 and 224 associated eSNPs were identified in the BC1_LD, BC1_DU and BC1_PI animals, respectively (Supplementary Table S2). A total of 26 eQTLs were found in BC1_LD located on SSC1-SSC11, SSC13 and SSC16. In BC1_DU, 32 eQTLs were detected on SSC1-SSC4, SSC6, SSC7, SSC9, SSC11-SSC13, SSC15, SSC17, and SSC18, and the 25 eQTLs found in BC1_PI were located on SSC1-SSC3, SSC6-SSC10, SSC12, SSC14 and SSC16-SSC18, and are represented in Figure 4 (Supplementary Table S3).

<u>*Cis*-eQTLs:</u>

The *cis*-eQTL regions of *ACSM5* and *IGF2* genes, on SSC3 and SSC2 respectively, appeared segregating in all three backcrosses, which suggest that the Iberian boars and the three founder maternal breeds have different allelic frequencies for the polymorphisms regulating *in cis* the expression of these genes.

The *ACSM5.P* polymorphism was segregating at low frequencies, being the *ACSM5.P A* allele frequency of 0.22 in BC1_LD, 0.09 in BC1_DU and 0.10 in BC1_PI. In the BC1_LD the *ACSM5.P* SNP was the most significant polymorphism associated with the differences in the mRNA level of *ACSM5*. However, in BC1_DU *rs81327383* was the most significantly associated SNPs (p-value=2.02x10^{-12}) with *ACSM5* mRNA expression although the *ACSM5.P* polymorphism was also significant (p-value=3.44x10^{-09}). In

BC1_PI, *rs81475068*, *rs81278505* and *ACSM5.P* polymorphisms were located on SSC3 and spanning 0.17 Mb (2.39-2.56 Mb) and were the most significant associated SNPs with *ACSM5* gene expression (*p*-value=7.32x10^{-09}). Hence, the lack of allele segregation or the presence of other proximal genetic variants could be involved in these gene expression changes.

In a previous work performed only in BC1_LD animals, the *cis*-eQTL for the muscle *IGF2* gene expression was identified, but the *IGF2*:*g.3072G>A* polymorphism was not the most significant associated SNP[17]. In the present work, r*s81322199* was located on SSC2 at 3.68 Mb and was the most significantly associated SNP in BC1_LD (*p*-value=1.45x10^{-15}), explaining the 42% of the phenotypic variance. In addition, the *IGF2g.3072G>A* polymorphism was significantly associated (*p*-value=3.03x10^{-07}) and explained the 22% of the *IGF2* mRNA variation. This result may be explained by the low number of homozygous AA animals, being 0.2 the allele frequency of the *IGF2:g.3072A* allele. On the other hand, the *IGF2g.3072G>A* polymorphism was the most significantly associated SNP with *IGF2* gene expression in BC1_DU and BC1_PI, explaining in both cases a high proportion of the gene expression variance, 58% and 92% respectively. In BC1_DU other genomic regions seem to be also associated with the *IGF2* gene expression differences, as the eQTL located in the 107.4-110.8 Mb genomic region of SSC4.

Two more *cis*-eQTLs were identified only in the BC1_LD population for *MGLL* and *NCOA2* gene expression. The *MGLL* eQTL was previously described in the same backcross[17]. The SSC4 *cis*-eQTL for *NCOA2* gene expression presented four significant associated SNPs, being the *rs80803396* the SNP showing the strongest signal (*p*-value=2.32x10^{-06}). Discrepancies between our results and the work of Puig-Oliveras *et al*. (2016)[17] may be explained by the different genome assemblies used between both works, being *Sscrofa* 10.2 genome assembly in the previous work and *Sscrofa* 11.1 in the present one.

Hotspots identified in t*rans*-eQTLs regions

All the *trans*-eQTLs intervals, eSNPs and annotated candidate genes are shown in the supplementary table S1, but only eQTL hotspots are discussed in detail

(Supplementary Table S4). In BC1_DU, new *trans*-eQTLs were identified for *ACAA2* (SSC1), *ACSM5* (SSC1, SSC2, SSC4, SSC6, SSC7, SSC11, SSC12, SSC13, and SSC18), *CREG1* (SSC1), *DGAT2* (SSC2), *ETS1* (SSC9), I*GF2* (SSC4), *LPIN1* (SSC4, SSC7, and SSC15), *NCOA1* (SSC1), *NCOA6* (SSC1), *PDHX* (SSC1), *PPARA* (SSC2, SSC15, and SSC17), *PRKAA1* (SSC1), and *PXMP3* (SSC4) genes. In BC1_LD additional *trans*-eQTLs were found for *ACSM5* (SSC1, SSC6, SSC8, and SSC10), *MLXIPL* (SSC2, SSC9, and SSC13), *CREG1* (SSC2), *DGAT2* (SSC2, SSC7, and SSC9), *FOS* (SSC11), *HIF1AN* (SSC2, SSC5, and SSC7), *MGLL* (SSC9), *PIK3R1* (SSC16), *PPARG* (SSC2), *PPARGC1A* (SSC2), and *SCD* (SSC2) genes. Finally, new *trans*-eQTLs in BC1_PI were detected for *ACSM5* (SSC1, SSC8, SSC12, SSC14, and SSC16), *ACSS2* (SSC7 and SSC18), *DGAT2* (SSC12 and SSC16), *HIF1AN* (SSC6 and SSC9), *LXRA* (SSC2), *PPARG* (SSC7, SSC10, SSC14, SSC16, and SSC17), *PPARGC1A* (SSC2, SSC6, SSC7, and SSC17), and *SCD* (SSC17) genes (Figure 4).

We only observed three common *trans*-eQTL regions in the 3BCs study, suggesting the presence of different regulatory mechanisms or frequencies according to breed. Overall, the *trans*-eQTL regions manifested that the expression of the genes related to lipid metabolism is regulated in a complex way.

In addition, six hotspots regions, two in each backcross, regulating the expression of several genes were detected.

In BC1_LD animals a *trans*-eQTL hotspot located on SSC2 and spanning 8.7 Mb (119.9-128.7 Mb) was associated with the expression of seven genes: *HIF1AN*, *CREG1*, *MLXIPL*, *DGAT2*, *PPARG*, *PPARGC1A,* and *SCD*. After gene annotation of this region no candidate *trans*-acting regulators modulating the expression of genes on the SSC2 hotspot were found. However, the transcription factor 7 (*TCF7*) gene was annotated in the *CREG1* eQTL region because it was six Mb longer (119.9-136.2 Mb) than the others. *TCF7* and its family member transcription factor 7 like 2 (*TCF7L2*) have been associated with diabetes in humans[37]. In addition, *TCF7L2* has been described as an indirect regulator of *PPARD* during adipogenesis[38]. In addition, to evaluate potential functional interactions and the co-expression pattern of genes on the SSC2 hotspot, GeneMANIA and PCIT co-expression network analysis were done (Figure 6). Interactions between *DGAT2*, *PPARG*, *PPARGC1A* and *SCD* were found with GeneMANIA (Figure 8A). In

general, meaningful gene-gene interactions were shown by PCIT algorithm (Figure 8B), reinforcing the presence of a common regulatory factor modulating the expression of SSC2 hotspot genes. However, lower correlations were observed for the *CREG1* gene, suggesting the presence of an independent regulatory factor modulating its expression. This result is in accordance with the proposal of the *TCF7* as a candidate gene of this region, although further validations are needed. Furthermore, *HIF1AN* presented negative and moderate correlations with *DGAT2*, *MLXIPL*, *PPARG*, and *SCD*, suggesting an opposite regulatory effect for this gene. *HIF1AN* gene is involved in fatty acid β-oxidation[39–41], while *DGAT2*, *PPARG*, *MLXIPL* and *SCD* genes are related to *de novo* lipogenesis, triacylglycerol synthesis and adipogenesis[16,42–44].

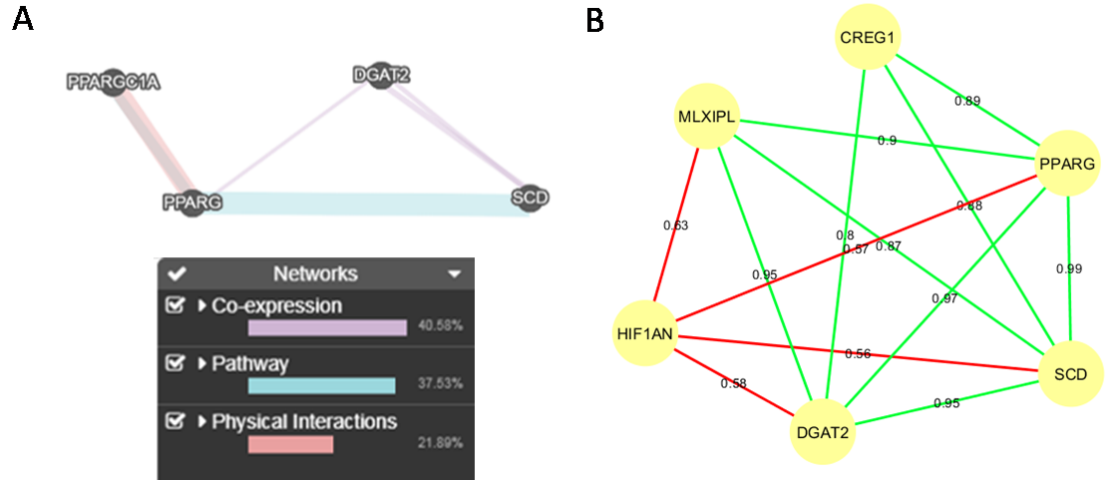


Figure 8: A) GeneMANIA analysis between SSC2 hotspot genes. B) Co-expression network using the PCIT algorithm within the genes associated with the BC1_LD *trans*-eQTL hotspot region on SSC2. Red and green lines indicate negative and positive correlations respectively.

The strong correlation for *SCD*, *PPARG* and *DGAT2* identified in the gene co-expression network in 3BCs, and with *MLXIPL* and *CREG1* have been found associated altogether with the *trans*-eQTL hotspot on SSC2 in the BC1_LD study but not in the other two backcrosses (BC1_DU and BC1_PI).

The region spanning 3.5 Mb on SSC7 (62.4-65.9 Mb) presented significant associations with the *HIF1AN* and *DGAT2* gene expression. The nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor alpha (*NFKBIA*) gene was mapped in this

region. It is a transcription factor involved in immune response, but also plays a direct role in adipogenesis and fat accumulation[45,46]. *NFKBIA* was found differentially expressed in different development stages and muscles between Iberian and Iberian x Duroc pigs, suggesting that it is a molecular regulator of metabolism[32]. An experimental interaction between *HIF1AN* and *NFKBIA* was identified by GeneMANIA and String programs, but no information about *DGAT2* interactions was found, so further validation will be needed to corroborate our results. Hence, we can suggest that *NFKBIA* is involved in muscle lipid metabolism, being an interesting candidate gene to explain the differences in the expression of two genes associated with the SSC7 hotspot in BC1_LD animals.

In the BC1_DU animals study two *trans*-eQTL hotspot regions were found on SSC1, spanning 6 Mb (180.6-203.6 Mb), and on SSC15, spanning 0.3 Mb (103.7-104 Mb). The SSC1 region showed significant associations with the expression of the *ACAA2*, *CREG1*, *NCOA1*, *NCOA6*, *PDHX* and *PRKAA1* genes. The perilipin 2 (*PLIN2*) gene was mapped in this region but was only annotated as a candidate gene for *ACAA2*, *NCOA1*, *NCOA6* and *PDHX*. *PLIN2* was reported to be involved in the uptake and storage of FAs in human skeletal muscle[47]. Studies in pigs described that a higher *PLIN2* gene expression was associated with a higher IMF content in muscle[48,49]. In order to deep in the study of the genes regulated by the same eQTL on SSC1, the PCIT algorithm was used to build a co-expression network. Moderate to high positive correlations, from 0.15 to 0.78, were observed among the genes regulated by the same eQTL (Figure 9). Lower correlations were observed for *ACAA2*, a gene encoding an enzyme that catalyzes the last step in mitochondrial fatty acid β-oxidation[50], suggesting the presence of another genetic factor regulating its expression. In addition, moderate correlations were found for the rest of the hotspot genes, mainly related to transcriptional regulation and control.

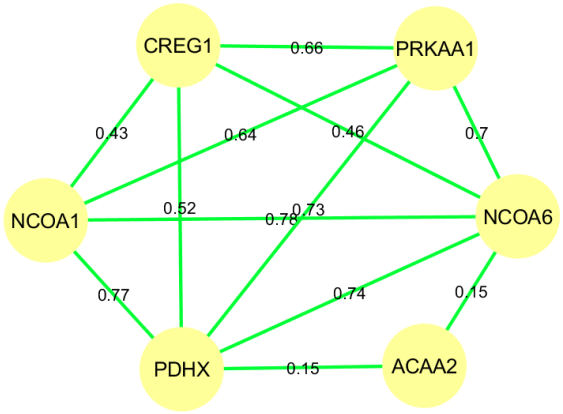


Figure 9: Co-expression network for genes associated with the BC1_DU *trans*-eQTL hotspot on SSC1 using the PCIT algorithm.

Notably, the second group of genes identified in the gene co-expression network in 3BCs, which showed strong correlations for *PRKAA1*, *PDHX*, *NCOA1* and *NCOA2* among others, coincides with the previously observed SSC1 *trans*-eQTL hotspot in BC1_DU study, but not in the other two backcrosses (BC1_LD and BC1_PI).

LPIN1 and *PPARA* genes were significantly associated with the SSC15 hotspot region and showed a moderate correlation value (*LPIN1-PPARA,* r=0.59 *p*-value=4.97x10^{-13}). In this region was mapped a key mitochondrial enzyme for fatty acid oxidation, *AOX1* gene. It has been reported to be associated with FA oxidation in mice adipocytes[51] and meat quality traits and with muscle development in cattle[52].

Regarding BC1_PI population, two *trans*-eQTL hotspots regions on SSC7 and SSC17 were observed. The first region, spanning 8.1 Mb on SSC7 (100.1-108.2 Mb), showed a significant association with *PPARG* and *PPARGC1A* gene expression. *DIO2* gene was mapped in the SSC7 *trans*-eQTL region as a potential candidate gene for lipid metabolism. It has been selected as a muscle candidate gene in an obesity resistance study since it presented differences between lean and fat mouse lines[53]. *DIO2* converts prohormone thyroxine (T4) to the active hormone triiodothyronine (T3), which binds to tyroid hormone receptors (TR). TR heterodimerize with RXR and can compete with PPAR for that binding site affecting gene control and regulation[54]. Hence, *DIO2* may be an indirect regulator of SSC7 hotspot genes.

The second region located on SSC17 and spanning 12.6 Mb (29.2-41.8 Mb), presented a significant association with *PPARG*, *PPARGC1A* and *SCD*. Three genes were mapped for the SSC17 hotspot: *RBL1*, *FOXA2* and *E2F1*. *RBL1* gene has been associated with the whole body fat metabolism and determines the oxidative state of muscle in mice[55]. *FOXA2* has been described as a transcription factor of several genes involved in the insulin pathway in liver[56], but no studies in muscle tissue were found. It was reported that *E2F1* is required for in vivo skeletal muscle regeneration in mouse[57] and showed high gene expression levels in Pietrain pigs with high muscle content[58]. Interactions were found between the genes associated with the hotspot (*PPARG*, *PPARGC1A* and *SCD)* and between the *E2F1*, *RBL1* and *PPARG* genes using GeneMANIA and String. *RBL1* and *E2F1* were selected as promising candidate genes for lipid metabolism in pigs, but further validations are needed to assess the effect of *FOXA2* in muscle tissue.

Conclusions

In the present study, we identified genetic variants associated with the gene expression of six lipid-related genes in muscle. Both *IGF2*:g.3072G>A and *ACSM5.P* polymorphisms were described as major regulators of *IGF2* and *ACSM5* gene expression levels respectively, of different genetic backgrounds, while different *trans*-eQTL hotspot regions were found in each backcross suggesting the presence of different regulatory mechanisms depending on the breed. In addition, sex-dimorphism and breed effects were found for the expression levels of most of the genes analysed and two groups of co-expressed genes were identified. Our results increase the knowledge of the genetic basis of gene expression regulation in muscle lipid metabolism. Overall, expression of genes related to lipid metabolism is regulated in a complex way.

Material and Methods

Animal material

Three different experimental backcrosses, BC1_LD (25% Iberian and 75% Landrace), BC1_DU (25% Iberian and 75% Duroc) and BC1_PI (25% Iberian and 75% Pietrain), belonging to a total of 355 animals were studied (called 3BCs): 114 BC1_LD, 122 BC1_DU and 119 BC1_PI.

All animals were maintained under the same intensive conditions and fed *ad libitum* with a cereal-based commercial diet and slaughtered in a commercial abattoir following institutional and national guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA – Institut de Recerca i Tecnologia Agroalimentàries). In addition, animal care and procedures were carried out according to the Spanish Policy for Animal Protection RD1201/05 and the European Union Directive 86/609 about the protection of animals used in experimentation.

LD samples were collected at slaughterhouse in liquid nitrogen and stored at -80°C until analysis. Genomic DNA was extracted from diaphragm tissue using the phenol-chloroform method[59].

Genotyping

Animals from BC1_LD and BC1_PI were genotyped using the Porcine SNP60K BeadChip (Illumina, San Diego, USA) and BC1_DU animals were genotyped using the Axiom Porcine Genotyping Array (Affymetrix). Only SNPs that mapped against the *Sscrofa* 11.1 assembly and were common to both arrays were selected. Markers that showed a minor allele frequency (MAF) lower than 5% and SNPs with more than 5% of missing genotypes were removed with Plink software[60]. Moreover, based on the information in the prior BC1_LD study[17], two additional SNPs were genotyped: *ACSM5* (*rs331702081*) and *IGF2* (*IGF2*:g.3072G>A), in the BC1_DU and BC1_PI populations, following the previously described protocols[17,22]. Finally, a total of 38.426 SNPs distributed along all chromosomes, including *rs331702081* and *IGF2*:g.3072G>A polymorphisms, were used for association studies.

Gene expression

Total RNA was obtained from the LD muscle of 355 animals using the RiboPure kit (Ambion), following the producer's recommendations**.** RNA quantification and purity was performed with a NanoDrop ND-1000 spectrophotometer (NanoDrop products) and RNA integrity was assessed by Agilent Bioanalyzer-2100 (Agilent Technologies). The RNA was reverse-transcribed into cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems), following the manufacturer's instructions.

Gene expression was analyzed in 48 genes, of which 45 were target genes and 3 were candidate reference genes (*ACTB*, *HPRT1* and *TBP*), by quantitative real time-PCR (qPCR). Selection of target genes related to lipid metabolism as well as primer design details and sequences was described in Puig-Oliveras *et al*. (2016)[17]. Gene expression quantification was performed in a 48.48 Microfluidic Dynamic Array IFC Chip (Fluidigm) in a BioMark System following a previously described protocol[61]. Data was collected using Fluidigm Real-Time PCR analysis software 3.0.2 (Fluidigm) and analyses were done with DAG Expression software 1.0.4.11[62], applying the relative standard method curve. In order to normalize the expression levels of target genes, *ACTB* and *TBP* were used as the most stable reference genes, and *HPRT1* was discarded. The normalized quantity (NQ)[62] values of each sample and assay were used to compare the expression data among animals. Normalization of data was checked through Shapiro-Wilk test in R[63], and $\log_2$ transformation of the NQ value was applied if necessary. Sex and breed effects were tested by using a linear model (lm) in R[63].

Genome-wide association analysis for gene expression

Genomic association studies between 45 gene expression values and common SNPs genotypes (eGWAS) were performed through a linear mixed model using GEMMA software[64]:

$$\mathbf{y} = \mathbf{W\alpha} + \mathbf{x}\beta + \mathbf{u} + \boldsymbol{\varepsilon};\ \mathbf{u} \sim MVN_n(0, \lambda\tau^{-1}\mathbf{K}),\ \boldsymbol{\varepsilon} \sim MVN_n(0, \tau^{-1}\mathbf{I}_n),$$

in which: y was the vector of phenotypes for n individuals; **W** is a matrix nxc of covariables (fixed effects) that includes sex (2 levels), backcross (3 levels) and batch (9

levels); **α** is a c vector with corresponding coefficients, including the intercept; **x** is a n vector with the marker genotypes; β is the size of the marker effect, **u** is an n vector of random effects (additive genetic effects), **ε** is an n vector of errors. The random effects vector is assumed to follow a normal multivariate n-dimensional distribution (MVN_n) where τ^{-1} is the variance of residual errors; λ is the quotient between the two components of variance; K is an nxn matrix of kin calculated from the SNPs. The vector of errors is assumed to follow a distribution MVN_n, where $\mathbf{I}_n$ is an nxn identity matrix.
GEMMA software calculates from the Wald statistical test the *p*-value for each SNP comparing the null hypothesis that the SNP has no effect versus the alternative hypothesis that the SNP effect is different from zero. The FDR (False Discovery Rate) method of Benjamini and Hochberg[65] was used for the correction of multiple tests with the function *p.adjust* of R software.

Gene annotation

Significant associated SNPs were mapped in the *Sscrofa* 11.1 assembly and were annotated with the Ensembl Genes 91 Database using VEP software[66]. BioMart software[67] was used to annotate genomic eQTL intervals considering ±1 Mb around the candidate chromosomal regions. In the three studied BCs study only eQTL intervals containing 2 or more SNPs were annotated, whereas in the individual backcross GWAS annotation was done for eQTL intervals containing 3 or more SNPs.

The identified SNPs were classified depending on their location, as *cis* if the SNPs were located within 1 Mb of the analyzed gene and as *trans* if the SNPs were located elsewhere in the genome. The number of significant SNPs belonging to the same interval was considered among associated SNPs less than 10 Mb apart.

Co-expression and functional analysis

The PCIT algorithm was used to calculate weighted gene co-expression networks, through the implementation of first-order partial correlations coefficients combined with information theory approach, in order to identify principal interactions between genes[68,69]. Only the significant interactions between genes were considered for further steps. Networks were represented with CentiScaPe Cytoscape plug-in[70].

Ingenuity Pathway Analysis software (IPA; Ingenuity Systems) and the *Core Analysis* function was used to perform functional analysis of genes mapped in the different intervals and for data interpretation in the context of biological processes, pathways and networks. In addition, the iRegulon v1.3. Cytoscape plug-in[71] was used to identify transcription factor binding sites *in silico*. ClueGO plug-in[72] was used to integrate and cluster the genes regarding their Gene Ontology and KEGG pathway. Finally, GeneMANIA[73] and String[74] were used to evaluate the functional interaction and networks among genes proteins, respectively.

Acknowledgments

We wish to thank all of the members of the INIA, IRTA, and UAB institutions who contributed to the generation of the animal material used in this work.

Author contribution statement

JMF and AIF conceived and designed the experiments; JMF was the principal investigator of the project; AIF and JMF collected the animal samples; LCM and DCP performed the pig genomic DNA extraction; AC genotyped the samples; LCM, MB and AC designed and performed the gene expression studies; LCM and JMF performed the genome-wide association studies; LCM, MB and JMF wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Spanish Ministerio de Economía y Competitividad (MINECO) and the Fondo Europeo de Desarrollo Regional (FEDER) with project references: AGL2014-56369-C2 and AGL2017-82641-R. L. Criado-Mesas was funded with an FPI grant from the AGL2014-56369-C2 project. M. Ballester was financially supported by a "Ramón y Cajal" contract (RYC-2013-12573) from the Spanish Ministerio de Economía y Competitividad. D. Crespo-Piazuelo was funded by a "Formació i Contractació de Personal Investigador Novell" (FI-DGR) Ph.D grant from the Generalitat de Catalunya (ECO/1788/2014). We acknowledge the support of the Spanish Ministerio de Economía y Competitividad for the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533) to the Centre for Research in Agricultural Genomics and the CERCA Programme / Generalitat de Catalunya.

References

1. Wood, J. D. *et al.* Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* **78**, 343–358 (2008).
2. Simopoulos, A. P. The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomed. Pharmacother.* **56**, 365–379 (2002).
3. Lopez-Bote, C. J. Sustained utilization of the Iberian pig breed. *Meat Sci.* **49**, (1998).
4. Ventanas, S. & Al, E. Quality traits in muscle biceps femoris and back-fat from purebred Iberian and reciprocal Iberian × Duroc crossbred pigs. *Meat Sci.* **73**, 651–659 (2006).
5. Wood, J. D. *et al.* Effects of fatty acids on meat quality: A review. *Meat Sci.* **66**, 21–32 (2004).
6. Casellas, J. *et al.* Bayes factor analyses of heritability for serum and muscle lipid traits in Duroc pigs. *J. Anim. Sci.* **88**, 2246–2254 (2010).
7. Ntawubizi, M. *et al.* Genetic parameters for intramuscular fatty acid composition and metabolism in pigs. *J. Anim. Sci.* **88**, 1286–1294 (2014).

8. Sellier, P. & Monin, G. Genetics of Pig Meat Quality: a Review. *J. Muscle Foods* **5**, 187–219 (1994).
9. Ramayo-Caldas, Y. *et al.* Genome-wide association study for intramuscular fatty acid composition in an Iberian x Landrace cross. *J. Anim. Sci.* **90**, 2883–2893 (2012).
10. Muñoz, M. *et al.* Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. *BMC Genomics* **14**, 845 (2013).
11. Revilla, M. *et al.* New insight into the SSC8 genetic determination of fatty acid composition in pigs. *Genet. Sel. Evol.* **46**, 1–10 (2014).
12. Ayuso, M. *et al.* Comparative analysis of muscle transcriptome between pig genotypes identifies genes and regulatory mechanisms associated to growth, Fatness and metabolism. *PLoS One* **10**, 1–33 (2015).
13. Corominas, J. *et al.* Polymorphism in the ELOVL6 Gene Is Associated with a Major QTL Effect on Fatty Acid Composition in Pigs. *PLoS One* **8**, 1–12 (2013).
14. Yang, B. *et al.* Genome-Wide Association Analyses for Fatty Acid Composition in Porcine Muscle and Abdominal Fat Tissues. **8**, (2013).
15. Hausman, G. J. *et al.* Board-invited review: The biology and regulation of preadipocytes and adipocytes in meat animals. *J. Anim. Sci.* **87**, 1218–1246 (2009).
16. Puig-Oliveras, A. *et al.* Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. *PLoS One* **9**, (2014).
17. Puig-Oliveras, A. *et al.* Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci. Rep.* **6**, 31803 (2016).
18. Ballester, M. *et al.* Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs. *Sci. Rep.* **7**, 46539 (2017).
19. Revilla, M. *et al.* Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions. *Sci. Rep.* **8**, 1–13 (2018).
20. Zhang, Y. *et al.* Transcriptional profiling of human liver identifies sex-biased

genes associated with polygenic dyslipidemia and coronary artery disease. *PLoS One* **6**, (2011).

21. Liu, D. *et al.* Skeletal muscle gene expression in response to resistance exercise: sex specific regulation. *BMC Genomics* **11**, 659 (2010).
22. Van Laere, A.-S. *et al.* A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003).
23. Varlamov, O., Bethea, C. L. & Roberts, C. T. Sex-specific differences in lipid and glucose metabolism. *Front. Endocrinol. (Lausanne).* **5**, 1–7 (2014).
24. Zhang, X. & Lerman, L. O. Investigating the Metabolic Syndrome. *Toxicol. Pathol.* **44**, 358–366 (2016).
25. Eberlé, D., Hegarty, B., Bossard, P., Ferré, P. & Foufelle, F. SREBP transcription factors: Master regulators of lipid homeostasis. *Biochimie* **86**, 839–848 (2004).
26. Benítez, R. *et al.* Modulatory effects of breed, feeding status, and diet on adipogenic, lipogenic, and lipolytic gene expression in growing iberian and duroc pigs. *Int. J. Mol. Sci.* **19**, 1–20 (2018).
27. Aslan, O. *et al.* Variation in the IGF2 gene promoter region is associated with intramuscular fat content in porcine skeletal muscle. *Mol. Biol. Rep.* **39**, 4101–4110 (2012).
28. Markljung, E. *et al.* ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biol.* **7**, (2009).
29. Criado-Mesas, L. *et al.* Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition. *PLoS One* **14**, e0220708 (2019).
30. Chang, Y. H., Huang, C. N. & Shiau, M. Y. Association of IL-4 receptor gene polymorphisms with high density lipoprotein cholesterol. *Cytokine* **59**, 309–312 (2012).
31. Barthel, A., Schmoll, D. & Unterman, T. G. FoxO proteins in insulin action and metabolism. *Trends Endocrinol. Metab.* **16**, 183–189 (2005).
32. Ayuso, M. *et al.* Developmental stage, muscle and genetic type modify muscle transcriptome in pigs: Effects on gene expression and regulatory factors involved in growth and metabolism. *PLoS One* **11**, 1–33 (2016).
33. Judge, S. M. *et al.* Genome-wide identification of FoxO-dependent gene

networks in skeletal muscle during C26 cancer cachexia. 1–17 (2014).

34. Hu, X. *et al.* Molecular cloning, expression pattern analysis of porcine Rb1 gene and its regulatory roles during primary dedifferentiated fat cells adipogenic differentiation. *Gen. Comp. Endocrinol.* **214**, 77–86 (2015).
35. Huang, P. *et al.* The possible role of complete loss of myostatin in limiting excessive proliferation of muscle cells (C2C12) via activation of microRNAs. *Int. J. Mol. Sci.* **20**, 643 (2019).
36. Zhang, B. *et al.* The association of NR1H3 gene with lipid deposition in the pig. *Lipids Health Dis.* **15**, 4–11 (2016).
37. Elbein, S. C., Das, S. K., Hallman, D. M., Hanis, C. L. & Hasstedt, S. J. Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American diabetes association GENNID (Genetics of NIDDM) study cohort. *Diabetes* **58**, 268–274 (2009).
38. Cristancho, A. G. *et al.* Repressor transcription factor 7-like 1 promotes adipogenic competency in precursor cells. *Proc. Natl. Acad. Sci.* **108**, 16271–16276 (2011).
39. Krishnan, J. *et al.* Dietary obesity-associated hif1α activation in adipocytes restricts fatty acid oxidation and energy expenditure via suppression of the Sirt2-NAD+ system. *Genes Dev.* **26**, 259–270 (2012).
40. Mylonis, I., Simos, G. & Paraskeva, E. Hypoxia-Inducible Factors and the Regulation of Lipid Metabolism. *Cells* **8**, 214 (2019).
41. Knutti, D. & Kralli, A. PGC-1, a versatile coactivator. *Trends Endocrinol. Metab.* **12**, 360–365 (2001).
42. Puig-Oliveras, A. *et al.* A co-association network analysis of the genetic determination of pig conformation, growth and fatness. *PLoS One* **9**, 1–20 (2014).
43. Ahmadian, M. *et al.* Pparγ signaling and metabolism: The good, the bad and the future. *Nat. Med.* **19**, 557–566 (2013).
44. Jiang, Z. *et al.* Significant associations of stearoyl-CoA desaturase (SCD1) gene with fat deposition and composition in skeletal muscle. *Int. J. Biol. Sci.* **4**, 345–351 (2008).
45. Yu, K. *et al.* Activating transcription factor 4 regulates adipocyte differentiation

via altering the coordinate expression of CCATT/enhancer binding protein β and peroxisome proliferator-activated receptor γ. *FEBS J.* **281**, 2399–2409 (2014).

46. Ren, W. *et al.* CCAAT/enhancer-binding protein α is a crucial regulator of human fat mass and obesity associated gene transcription and expression. *Biomed Res. Int.* **2014**, (2014).
47. Bickel, P. E., Tansey, J. T. & Welte, M. A. PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1791**, 419–440 (2009).
48. Davoli, R. *et al.* New SNP of the porcine Perilipin 2 (PLIN2) gene, association with carcass traits and expression analysis in skeletal muscle. *Mol. Biol. Rep.* **38**, 1575–1583 (2011).
49. Gandolfi, G. *et al.* Perilipin 1 and perilipin 2 protein localization and gene expression study in skeletal muscles of European cross-breed pigs with different intramuscular fat contents. *Meat Sci.* **88**, 631–637 (2011).
50. Bartlett, K. & Eaton, S. Mitochondrial β-oxidation. *Eur. J. Biochem.* **271**, 462–469 (2004).
51. Gan, L., Liu, Z., Cao, W., Zhang, Z. & Sun, C. FABP4 reversed the regulation of leptin on mitochondrial fatty acid oxidation in mice adipocytes. *Sci. Rep.* **5**, 1–12 (2015).
52. Guillocheau, G. M. *et al.* Survey of allele specific expression in bovine muscle. *Sci. Rep.* **9**, 1–11 (2019).
53. Simončič, M. *et al.* Obesity resistant mechanisms in the Lean polygenic mouse model as indicated by liver transcriptome and expression of selected genes in skeletal muscle. *BMC Genomics* **12**, 1–12 (2011).
54. Sinha, R. A., Singh, B. K. & Yen, P. M. Thyroid hormone regulation of hepatic lipid and carbohydrate metabolism. *Trends Endocrinol. Metab.* **25**, 538–545 (2014).
55. Scimè, A. *et al.* Oxidative status of muscle is determined by p107 regulation of PGC-1α. *J. Cell Biol.* **190**, 651–662 (2010).
56. K.A., L. *et al.* Foxa2 regulates multiple pathways of insulin secretion. *J. Clin. Invest.* **114**, 512–520 (2004).
57. Yan, Z. *et al.* Highly coordinated gene regulation in mouse skeletal muscle regeneration. *J. Biol. Chem.* **278**, 8826–8836 (2003).

58. Fan, H. the Hydrogoniometer and Assessment of Gleno-Humeral Joint Motion. *Gene* **486**, 8–14 (2011).
59. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual 2nd edn. Cold Spring Harbor, N.Y.* doi:0167-7799(91)90068-S
60. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
61. Ramayo-Caldas, Y. *et al.* From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. *BMC Genomics* **15**, 232 (2014).
62. Ballester, M., Cordón, R. & Folch, J. M. DAG expression: High-throughput gene expression analysis of real-time PCR data using standard curves for relative quantification. *PLoS One* **8**, 8–12 (2013).
63. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2018).
64. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–4 (2012).
65. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
66. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
67. Smedley, D. *et al.* The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).
68. Reverter, A. & Chan, E. K. F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* **24**, 2491–2497 (2008).
69. Watson-Haigh, N. S., Kadarmideen, H. N. & Reverter, A. PCIT: An R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics* **26**, 411–413 (2009).
70. Scardoni, G., Petterlini, M. & Laudanna, C. Analyzing biological network parameters with CentiScaPe. *Bioinformatics* **25**, 2857–2859 (2009).
71. Janky, R. *et al.* iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput. Biol.* **10**, (2014).

72. Bindea, G. *et al.* ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
73. Warde-Farley, D. *et al.* The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, 214–220 (2010).
74. Morris, J. H. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2018).

Supplementary Tables

Supplementary Table S1: List of significant associated SNPs within eQTLs intervals for the 45-muscle gene expression study in 3BCs.

Supplementary Table S2: List of significant associated SNPs within eQTLs intervals for the 45-muscle gene expression study in each backcross independently.

Supplementary Table S3: Significant eQTLs found for the 45-muscle gene expression study in each backcross independently. Start and end positions refer to the eQTL interval and are based on *Sscrofa* 11.1 assembly. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval. For the *cis*-eQTLs regions only the analyzed gene was annotated as positional candidate gene.

Supplementary Table S4: Significant *trans*-eQTLs for the hotspot regions found in each backcross independently. Start and end positions refer to the eQTL interval and are based on *Sscrofa* 11.1 assembly. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval.

PAPER II

Analysis of porcine *IGF2* gene expression in adipose tissue and its effect on fatty acid composition

Lourdes Criado-Mesas[1]*, Maria Ballester[2], Daniel Crespo-Piazuelo[1,3], Anna Castelló[1,3], Rita Benítez[4], Ana Isabel Fernández[4] and Josep M. Folch[1,3]

[1]Departament de Genòmica Animal, Centre de Recerca en Agrigenòmica (CRAG), CSIC-IRTA-UAB-UB, Barcelona, Spain.

[2]Departament de Genètica i Millora Animal, Institut de Recerca y Tecnologia Agraroalimentàries (IRTA), Caldes de Montbui, Spain.

[3]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, UAB, Bellaterra, Spain.

[4]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain.

* Corresponding author

PLoS ONE (2019) 14(8): e0220708.
https://doi.org/10.1371/journal.pone.0220708

Abstract

IGF2:*g.3072G>A* polymorphism has been described as the causal mutation of a maternally imprinted QTL for muscle growth and fat deposition in pigs. The objective of the current work was to study the association between the *IGF2*:*g.3072G>A* polymorphism and the *IGF2* gene expression and its effect on fatty acid composition in adipose tissue in different pig genetic backgrounds.

A *cis*-eQTL region associated with the *IGF2* mRNA expression in adipose tissue was identified in an eGWAS with 355 animals. The *IGF2* gene was located in this genomic interval and *IGF2g.3072G>A* was the most significant SNP, explaining a 25% of the gene expression variance. Significant associations between *IGF2*:*g.3072G>A* polymorphism and oleic (C18:1(n-9); p-value=$4.18x10^{-07}$), hexadecanoic (C16:1(n-9); p-value=$4.04x10^{-07}$), linoleic (C18:2(n-6); p-value=$6.44x10^{-09}$), α-linoleic (C18:3(n-3); p-value=$3.30x10^{-06}$), arachidonic (C20:4(n-6); p-value=$9.82x10^{-08}$) FAs and the MUFA/PUFA ratio (p-value=$2.51x10^{-9}$) measured in backfat were identified. Animals carrying the *A* allele showed an increase in *IGF2* gene expression and higher PUFA and lower MUFA content. However, in additional studies was observed that there could be other proximal genetic variants affecting FA composition in adipose tissue.

Finally, no differences in the *IGF2* gene expression in adipose tissue were found between heterozygous animals classified according to the *IGF2:g.3072G>A* allele inherited from the father (A^PG^M or A^MG^P). However, pyrosequencing analysis revealed that there is imprinting of the *IGF2* gene in muscle and adipose tissues, with stronger differences among the paternally and maternally inherited alleles in muscle.

Our results suggested that *IGF2*:*g.3072G>A* polymorphism plays an important role in the regulation of *IGF2* gene expression and can be involved in the fatty acid composition in adipose tissue. In both cases, further studies are still needed to deepen the mechanism of regulation of *IGF2* gene expression in adipose tissue and the *IGF2* role in FA composition.

Introduction

Over the last few years there has been a highlighted interest in identifying genes that improve meat quality. The nutritional value of meat and its quality is determined by several factors, including the intra-muscular fat (IMF) content and its fatty acid (FA) composition. Fat tissue firmness, shelf life, flavour, tenderness and juiciness [5] are influenced by FA composition, which is also involved in both meat nutritional traits and common diseases such as obesity and diabetes [1].

A paternally expressed Quantitative Trait Locus (QTL) for muscle growth and backfat (BF) thickness was identified in pig chromosome 2 (SSC2), in a genomic region containing the *insulin-like growth factor 2* (*IGF2*) gene [75,76]. *IGF2* is a maternally imprinted gene which promotes growth and plays an important role in proliferation, differentiation and apoptosis of cells in different tissues [77,78]. Moreover, IGF2 dysfunctions are involved in metabolic disorders, such as diabetes and obesity among others [79]. Latterly, IGF2 has been proposed as a physiological regulator of preadipocyte growth, metabolism and body fat composition in humans [80,81], although regulation of the *IGF2* gene is still uncertain.

Some years later, the polymorphism *g.3072G>A* located in the intron 3 of the *IGF2* gene was described as the causal mutation for this QTL, which increases muscle growth and heart size and reduces subcutaneous fat deposition [22]. The mutation is located in a well-conserved CpG island that is hypomethylated in skeletal muscle and abrogates the binding site for ZBED6, a nuclear factor which repress *IGF2* transcription, leading to a 3-fold up-regulation of *IGF2* expression in skeletal muscle [28]. The *IGF2*:*g.3072G>A* polymorphism has been associated with *IGF2* expression in muscle, but not in liver [22] and adipose [82] tissues, indicating a tissue-dependent regulation of *IGF2* gene expression.

The causal mutation for this QTL is widespread in different breeds [83] and it contributes to the improvement of porcine production, explaining 15-30% of the phenotypic variation in muscle mass and 10-20% of the variation in BF thickness [75,76]. The effects of this mutation on several growth traits have also been identified in different populations. For example, in a Large White commercial population and in

an Iberian x Landrace F2 cross the *IGF2* polymorphism was associated with BF thickness, carcass weight, *longissimus* muscle area, ham weight and shoulder weight traits [84].

Furthermore, an association between the *IGF2* gene expression and the percentage of IMF, in which animals with a high gene expression presented greater IMF content in skeletal muscle, has been described [27]. Another study showed that the mutation has an effect on both carcass and ham conformation and they detected an increase in monounsaturated FA (MUFA) and a decrease in polyunsaturated FA (PUFA) content in hams of pigs carrying the *A* allele [85]. However, there is a lack of literature analysing the effect of the *IGF2*:*g.3072G>A* polymorphism on FA composition measured in adipose tissue.

The aim of this work was to study the association between the *IGF2*:*g.3072G>A* polymorphism and the *IGF2* gene expression in adipose tissue to better understand 1) the *IGF2* gene expression regulation in adipose tissue and 2) the effect of the *IGF2* gene on adipose tissue FA composition.

Material and methods

Animal material

A total of 355 animals belonging to different experimental backcrosses, BC1_LD (25% Iberian and 75% Landrace), BC1_DU (25% Iberian and 75% Duroc) and BC1_PI (25% Iberian and 75% Pietrain), were analyzed. This set of animals from three different backcrosses was named 3BCs. All animals were maintained under intensive conditions and feeding was *ad libitum* with a cereal-based commercial diet. Animal procedures were performed according to the Spanish Policy for Animal Protection RD1201/05, which meets the European Union Directive 86/609 about the protection of animals used in experimentation. The experimental protocol was approved by the Ethical Committee of IRTA (Institut de Recerca i Tecnologia Agroalimentàries).

BF samples were taken between the third and the fourth ribs, collected at slaughter in liquid nitrogen and stored at -80°C until analysis. Genomic DNA was extracted from diaphragm tissue samples using the phenol-chloroform method [59].

Phenotypic data

Composition of 17 FAs in the C:12 and C:22 range in BF adipose tissue was determined by gas chromatography of methyl esters [86]. Afterwards, the percentage of the content of each FA was calculated in addition to the overall percentage of saturated FAs (SFA), MUFA and PUFA. In addition, ratios of FA as indices for desaturation and elongation were determined. BF thickness was measured between the 4^{th} and the 5^{th} ribs.

Genotyping

Animals from BC1_LD and BC1_PI were genotyped with Porcine SNP60K BeadChip (Illumina, San Diego, USA) and BC1_DU animals with Axiom Porcine Genotyping Array (Affymetrix, Inc.).

Common Single Nucleotide Polymorphisms (SNPs) in both arrays were mapped against the *Sus scrofa 11.1* assembly and Plink software [60] was used afterwards to remove markers that showed a minor allele frequency (MAF) less than 5% and SNPs with more than 5% of missing genotypes. After filtering, a total of 38,424 SNPs were retained for association studies. In addition, the *IGF2*:*g.3072G>A* polymorphism was genotyped using a pyrosequencing protocol previously described [22] and a SNP located in the predicted 3' UTR region of the gene (*ENSSSCT00000039341.1:c.1469990C>T*) was genotyped using Taqman OpenArrayTM genotyping plates custom-designed in a QuantStudioTM 12K flex Real-Time PCR System (ThermoFisher Scientific).

Gene expression

Reverse transcription quantitative real time-PCR (RT-qPCR) was used to study *IGF2* gene expression in a total of 355 animals from BC1_LD (n=114), BC1_DU (n=122) and BC1_PI (n=119) in BF adipose tissue. In addition, the *Longissimus dorsi* (LD) muscle *IGF2* expression was analysed in 14 animals corresponding to BC1_LD (n=7) and BC1_DU (n=7). Total RNA was obtained using the RiboPure kit (Ambion), following the

producer'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). One microgram of total RNA was reverse-transcribed into cDNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) using random hexamer primers in 20 µl reactions, following the manufacturer's instructions. Minus reverse transcription polymerase controls were also included to test for residual genomic DNA amplification. Primers for *IGF2* and two reference genes, *actin beta* (*ACTB*) and *TATA box binding protein* (*TBP*) (S1 Table), were designed using PrimerExpress 2.0 software (Applied Biosystems) [17].

Gene expression quantification in BF adipose tissue samples was performed in a QuantStudio™ 12K Flex Real-Time PCRSystem (ThermoFisher Scientific) using a 384-well plate and each sample was analyzed per triplicate. PCR amplifications were done in a final volume of 15 µl, including: 7,5 µl of SYBR® Select Master Mix (ThermoFisher Scientific), 300 nM of each primer and 3,75 µL of a 1:25 cDNA dilution. The PCR thermal cycle was: 2 min at 50ºC, 10 min at 95 °C, 40 cycles of 15 sec at 95 °C and 1 min at 60 °C. Moreover, a melting profile (95 ºC for 15 sec, 60 ºC for 15 sec and a gradual increase in temperature with a ramp rate of 1% up to 95 ºC) was added following the thermal cycling protocol, to assess for the specificity of the reactions. RT-qPCR efficiency for each assay was controlled using relative standard curves generated from a pool of cDNA from all samples serially diluted 5 fold. Data was collected and analyzed using the ThermoFisher Cloud software 1.0 (Applied Biosystems) applying the $2^{-\Delta\Delta Ct}$ [87] method for relative quantification (RQ) and using the lowest expression sample as calibrator.

Gene expression quantification in LD samples was performed in a 48.48 Microfluidic Dynamic Array IFC Chip (Fluidigm) in a BioMark System following a previously described protocol [61]. Data was collected and analysed using Fluidigm Real-Time PCR analysis software 3.0.2 (Fluidigm) and DAG Expression software 1.0.4.11 [62] respectively, applying the relative standard method curve.

Normalization of data was checked through Shapiro-Wilk test in R (https://r-project.org/) and log2 transformation was applied. A linear model (lm) was used also in R for test sex and breed effects [63].

Differential allelic expression quantification by pyrosequencing

A subset of 14 animals were selected based on their deduced paternally-inherited alleles and complete linkage disequilibrium between the *IGF2:g.3072G>A* and the *IGF2* 3' UTR (*ENSSSCT00000039341.1:c.1469990C>T*) polymorphisms, being all heterozygous for both variants. Seven animals carried the paternally derived haplotype *IGF2:g.3072 A - ENSSSCT00000039341.1:c.1469990 C* and 7 animals the alternative paternally derived *IGF2:g.3072 G - ENSSSCT00000039341.1:c.1469990 T* haplotype. Hence, analysis of the allelic expression at *ENSSSCT00000039341.1:c.1469990C>T* variant allowed us to infer the relative expression of *IGF2:g.3072G>A* alleles. Pyrosequencing analyses were performed in both muscle and adipose tissues.

A 114-bp fragment of the 3'-UTR region of *IGF2* gene containing the *ENSSSCT00000039341.1:c.1469990C>T* polymorphism was amplyfied using the following primers: Forward primer 5'-CACGCTCGCAGCTCTCTT-3', Reverse primer 5'-[biotin]CCCCCAGAAAGCTCGGAG-3' and pyrosequenced with primer 5'-CTCGCAGCTCTCTTG-3'.

RNA samples were treated with the Turbo DNA-free kit (Invitrogen, ThermoFisher Scientific) following manufacturer's instructions before reverse transcription. Amplification of cDNA samples was done using the GC RICH PCR system (Roche). Reactions included 1.5 mM of MgCl2, 200 µM of dNTP, 0.3 µM of each primer, 1U of GC-rich enzyme mix, 0.5 M GC-rich resolution solution and 2 µl of cDNA diluted 1:2 in a final volume of 25 µl. The thermal profile was 95 °C for 3 min, 40 cycles at 95 °C for 30 sec, 58 °C for 30 sec and 72 °C for 45 sec for the first 10 cycles and 5 more sec for each cycle in addition and a final extension step of 7 min at 72 °C. We tested whether amplification of genomic DNA was circumvented by RNA treatment with DNase using RNA not reverse-transcribed as a template.

The biotinylated PCR products were checked in high resolution agarose gels and analysed by pyrosequencing at the Sequencing and Functional Genomics Service of the Instituto Aragonés de Ciencias de la Salud (IACSs) with a PSQ 96MA system equipment (Biotage). Pyrosequencing data were analysed and quantified using the AQ mode of PSQ 96MA 2.1. software (Pyrosequencing QIAGEN). Calibration samples were prepared

by mixing homozygous genomic DNA samples (TT or CC) at different proportions to check the precision of the assay in estimating allele-specific frequencies.

Genome-wide association analysis for adipose tissue *IGF2* gene expression

Genomic association studies between gene expression values of *IGF2* and SNPs genotypes (eGWAS) were performed through a linear mixed model using GEMMA software [64]:

$$y = W\alpha + x\beta + u + \varepsilon;\ u \sim MVN_n(0, \lambda\tau^{-1}K), \varepsilon \sim MVN_n(0, \tau^{-1}I_n),$$

in which: y was the vector of phenotypes for n individuals; W is a matrix n×c of covariates (fixed effects) that includes a column of ones, sex (2 levels), backcross (3 levels), and batch (9 levels); α is a c vector with corresponding coefficients, including the intercept; x is an n vector with the marker genotypes; β is the size of the marker effect, u is an n vector of random effects (additive genetic effects), ε is an n vector of errors. The random effects vector is assumed to follow a normal multivariate n-dimensional distribution (MVN_n) where τ^{-1} is the variance of residual errors; λ is the quotient between the two components of variance; K is an n×n matrix of kinship calculated from the autosomal SNPs. The vector of errors is assumed to follow a distribution MVN_n, where I_n is an n×n identity matrix.

GEMMA software calculates the Wald statistical test and the P-value for each SNP comparing the null hypothesis that the SNP has no effect versus the alternative hypothesis that the SNP effect is different from zero. The FDR (False Discovery Rate) method of Benjamini and Hochberg [65] was used for the correction of multiple tests with the *p.adjust* function of R.

Gene annotation

The significantly associated SNPs were mapped in the *S. scrofa 11.1* assembly and were annotated with the Ensembl Genes 91 Database using VEP software [66]. The genomic eQTL intervals considering ±1 Mb around the candidate chromosomal regions were annotated using BioMart software [67].

The SNPs identified were classified as *cis* when they were located within 1 Mb from the gene analysed and as *trans* when they were located elsewhere in the genome.

Significant SNPs located less than 10 Mb apart were considered as belonging to the same genomic interval.

Association analysis for adipose tissue fatty acid composition

The linear mixed model previously described for eGWAS, adding carcass weight as a covariate, was carried out to study the association among 2,431 SSC2 SNPs genotypes and FA composition measured in BF tissue in 341 animals using GEMMA software [64].

Correlation analyses were done to better understand the relationship between gene expression and phenotypes. Gene expression was corrected by sex (two levels), backcross (three levels), and batch (nine levels) effects, and the FA composition was adjusted for sex, backcross, batch, and carcass weight. The corrected values of FA composition and gene expression were used to obtain the Pearson pairwise correlations.

Imprinting analysis

Paternal allele of 355 animals was deduced from progenitor's genotypes. An imprinting model of *IGF2* expression in BF was tested using a linear model (*lm*) in R, adjusting for sex, backcross, and batch as fixed effects. A comparison between this model and additive model was performed. The same models were tested with FA composition.

Results and discussion

Genome wide association study of adipose tissue *IGF2* gene expression

An eGWAS was performed among the genotypes of 38,425 SNPs, including *IGF2*:*g.3072G>A*, and the *IGF2* mRNA expression values in BF adipose tissue of 355 animals from all three backcrosses (3BCs) (Fig 1).

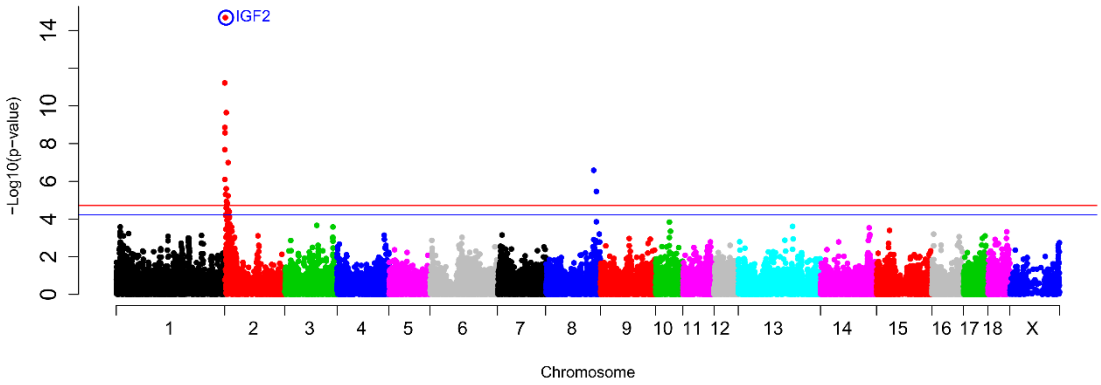


Fig 1. GWAS plot of *IGF2* gene expression in adipose tissue in the 3BCs animals. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

Two chromosomal regions (eQTLs) on SSC2 and SSC8 presented significant associations with the *IGF2* gene expression in adipose tissue using an additive model (Table 1).

Table 1. Significant eQTLs for adipose tissue *IGF2* gene expression in the 3BCs animals.

Re-gion	Chr	Start-End Positions	Size (Mb)	SNPs N	Start-End SNPs	Most Significant SNP	P-value	Type of eQTL	Candidate genes*
1	2	0,145,257-11,580,559	11.43	21	*rs81306755-rs81278022*	*IGF2*	2.07E-15	*cis/ trans*	*IGF2* and *SF1*
2	8	121,609,989-128,426,767	6.82	2	*rs80913047-rs81404614*	*rs80913047*	2.59E-07	*trans*	

Chromosomal location is based on *S.scrofa 11.1* assembly. Positions start and end refer to the eQTL interval. Gene annotation was performed considering one additional Mb

at the start and at the end of the eQTL interval. Number of SNPs corresponds to the SNPs within the eQTL interval. P-value corresponds to the most significant SNP. For the *cis*-eQTL only the analyzed gene was considered. *Genes with functions related to *IGF2*.

The SSC2 eQTL was divided in a *cis* and a *trans*-eQTL regions according to the distance from the *IGF2* gene. For the *cis*-eQTL region, where the *IGF2* gene was located, the *IGF2*:*g.3072G>A* mutation was the most associated SNP (p-value=$2.07x10^{-15}$). This result is in accordance with findings in muscle tissue, where I*GF2* mRNA expression was associated with this polymorphism [22] and suggests that it is also the causal mutation of *IGF2* gene expression in adipose tissue. A 25% of the phenotypic variance is explained by the *IGF2*:*g.3072G>A* polymorphism, indicating that other genetic variants and environmental factors are regulating *IGF2* gene expression in adipose tissue. In the same region, prior studies have reported the existence of *IGF2* antisense transcript in pigs and its coregulation with the *IGF2* gene in muscle and liver tissues. Furthermore, the antisense transcript was involved in the transcription regulation of *IGF2* promoters 2, 3 and 4 in post-natal muscle of animals carrying the *A* allele [88]. Therefore, it may be also involved in the regulation of *IGF2* in adipose tissue. In addition, *Splicing factor 1* (*SF1*) gene was mapped in the SSC2 *trans*-eQTL region (Table 1) and it was involved in the spliceosoma assembly and the alternative splicing which is an important mechanism for gene expression regulation [89].

Finally, *rs80913047* (p-value=$2.59x10^{-7}$) was the most significant associated SNP with BF *IGF2* expression on SSC8 but no candidate genes were annotated in this region.

eGWAS studies were also performed in animals of each backcross independently. In BC1_LD, no significant eQTLs regions were found (S1 Fig). This result is likely explained by the low number of animals with the *AA* genotype in the BC1_LD backcross (Table 2), being 0.2 the allele frequency of the *IGF2:g.3072A* allele.

Table 2. Summary of the number of animals used in this study.

		3BCs	BC1_LD	BC1_DU	BC1_PI
Sex	**Female**	186	65	63	58
	Male	169	49	59	61
Genotype	***AA***	131	4	61	66
	GA	148	39	56	53
	GG	76	71	5	0
Paternal Allele	***A***	145	23	61	61
	G	182	91	44	47

Number of animals are according to sex (n=355), the *IGF2g.3072G>A* polymorphism genotype (n=355) and the paternal allele genotype (n=355).

In addition, a previous work of our group performed in the *Longissimus dorsi* muscle of BC1_LD animals identified the *cis*-eQTL of the *IGF2* gene region, but the *IGF2*:*g.3072G>A* polymorphism was not the most significant SNP associated with the *IGF2* mRNA expression in muscle [17].

In contrast, the SSC2 and SSC8 eQTLs were also found in the BC1_DU backcross (Fig 2A), being the *rs81302016* SNP of the SSC8 the most significant associated SNP with the *IGF2* mRNA expression (p-value=2.17×10^{-7}). The *IGF2g.3072G>A* polymorphism was the most associated SNP on SSC2 (p-value=4.10×10^{-7}). Moreover, a proximal region located at 11.6 Mb of SSC2 showed a strong signal, being *rs81336616* (p-value=5.32×10^{-7}) the second most significant SNP in this region. The *IGF2:g.3072G>A* polymorphism explains a 24% of the phenotypic variance of adipose tissue *IGF2* gene expression in BC1_DU. A linear mixed model using the *IGF2*:*g.3072G>A* polymorphism as a fixed effect was analysed, showing no other additional eQTL on SSC2 for *IGF2* gene expression (S2 Fig). However, a second eQTL at 11.6 Mb of SSC2 may not be discarded due to the linkage disequilibrium observed between *IGF2*:*g.3072G>A* and *rs81336616*

SNPs in BC1_DU animals (R^2 = 0.464, D' = 0.693). Moreover, two additional *trans*-eQTLs regions were detected at SSC1 and SSC18 in the BC1_DU population (Table 3).

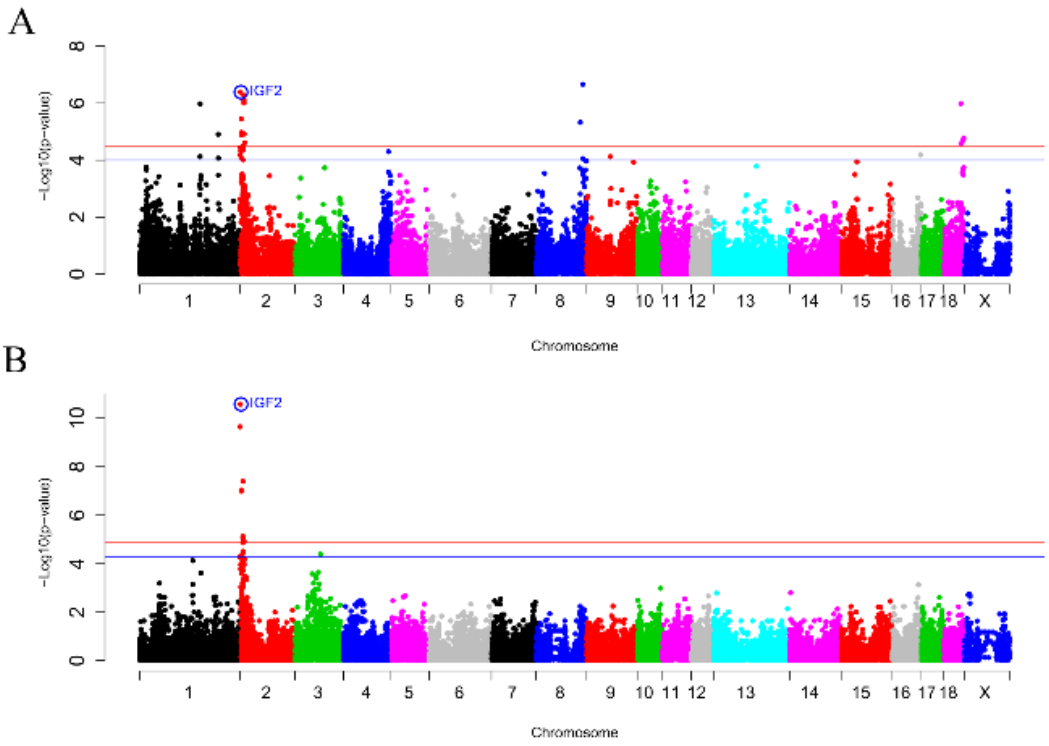


Fig 2. GWAS plot of adipose tissue *IGF2* gene expression in (A) BC1_DU and (B) BC1_PI. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value <0.1). The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

Table 3. Significant eQTLs for adipose tissue *IGF2* gene expression in BC1_DU and BC1_PI.

Re-gion	Chr	Start - End Pos.	Size (Mb)	SNPs N	Start-End SNPs	Most Significant SNP	*P*-value	Type of eQTL	Candidate genes*
1_DU	1	165.5 – 215.5	50	4	*rs81349445-rs80815028*	*rs80815896*	1.05E-06	*trans*	
2_DU	2	0.1 – 13.1	13	23	*rs81306755-rs332366314*	*IGF2g.3072G>A*	1.09E-06	*cis/ trans*	*IGF2* and *SF1*
3_DU	8	121.6 – 138.2	16.6	4	*rs80913047-rs81406196*	*rs81302016*	2.17E-07	*trans*	
4_DU	18	46.8 – 54.2	7.4	4	*rs81470467-rs81471417*	*rs81470467*	1.04E-06	*trans*	*IGFBP1* and *IGFBP3*
1_PI	2	0.07 – 11.4	11.3	19	*rs81341288-rs81361529*	*IGF2g.3072G>A*	2.64E-11	*cis/ trans*	*IGF2* and *SF1*

Regions corresponding to BC1_DU and BC1_PI are referenced as _DU or _PI respectively. Chromosomal location is based on *S. scrofa 11.1* assembly. Positions start and end (Mb) refer to the eQTL interval. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. Number of SNPs corresponds to the SNPs within the eQTL interval. P-value corresponds to the most significant SNP. For the *cis*-eQTL only the analyzed gene was considered. *Genes with functions related to *IGF2*.

Four significant associated SNPs were found in the SSC18 eQTL being *rs81470467* the most significant one (p-value=1.04x10^{-6}). Remarkably, two members of the IGF2 family were mapped in this region, the *insulin-like growth factor binding protein 1* and *3* (*IGFBP1* and *IGFBP3*), which are regulators of IGF activity, availability and tissue

distribution [90]. Specifically, *IGFBP1* gene was involved in obesity prevention and developing glucose intolerance and *IGFBP3* was described as an inducer of insulin resistance [91].

Finally, in the BC1_PI backcross only the SSC2 eQTL was found (Fig 2B), being the *IGF2*g:3072G>A polymorphism the most significantly associated SNP with *IGF2* expression in adipose tissue (p-value=2.64x10^{-11}) (Table 3). The 46% of the phenotypic variance was explained by the *IGF2*:*g.3072G>A* polymorphism, a higher proportion than in the other backcrosses.

When the *IGF2* expression values in adipose tissue of the 3BCs were classified according to the *IGF2*:*g.3072G>A* genotypes, animals with the *AA* genotype (mean=3.64, n=131) showed the highest mRNA expression with significant differences with *GA* (mean=2.53, n=148) and *GG* (mean=2.36, n=75) genotypes (*AA*-*GA*: p-value=1.78x10^{-13}, *AA*-*GG*: p-value=7.25x10^{-04}, *GA*-*GG*: p-value=9.9x10^{-02}) (Fig 3). Similar results were observed when the three backcrosses were analyzed separately.

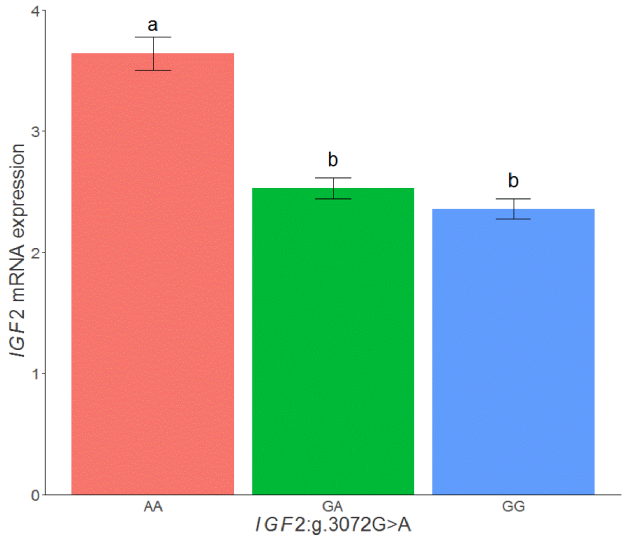


Fig 3. Plot of relative quantification of *IGF2* mRNA levels in adipose tissue of the 3BCs according to the *IGF2*:*g.3072G>A* SNP genotypes. Data represents means ± standard error of mean (SEM). Values with different superscript letters (a, b) indicate significant differences between groups (P-value <0.05).

Analysis of the imprinting effect on *IGF2* gene expression

To investigate the *IGF2* gene imprinting in adipose tissue, pyrosequencing analysis was performed in animals with known paternally-inherited alleles, both in muscle and adipose tissues (Fig 4). *IGF2* gene expression in both tissues was also obtained for these animals. As expected, in muscle the *A* allele percentage (from the sum of the two alleles) was higher (95.4%) in animals inheriting the *A* allele from his father than in animals inheriting the *G* allele (30.4%; p-value = 1.03×10^{-06}). In adipose tissue, the *A* allele percentage was also higher (79.3%) in animals inheriting the *A* allele from his father than in animals receiving the *G* allele (41.7%; p-value=0.002). According to these results, there is imprinting of the *IGF2* gene in both tissues, although stronger differences among the paternally and maternally inherited alleles were observed in muscle.

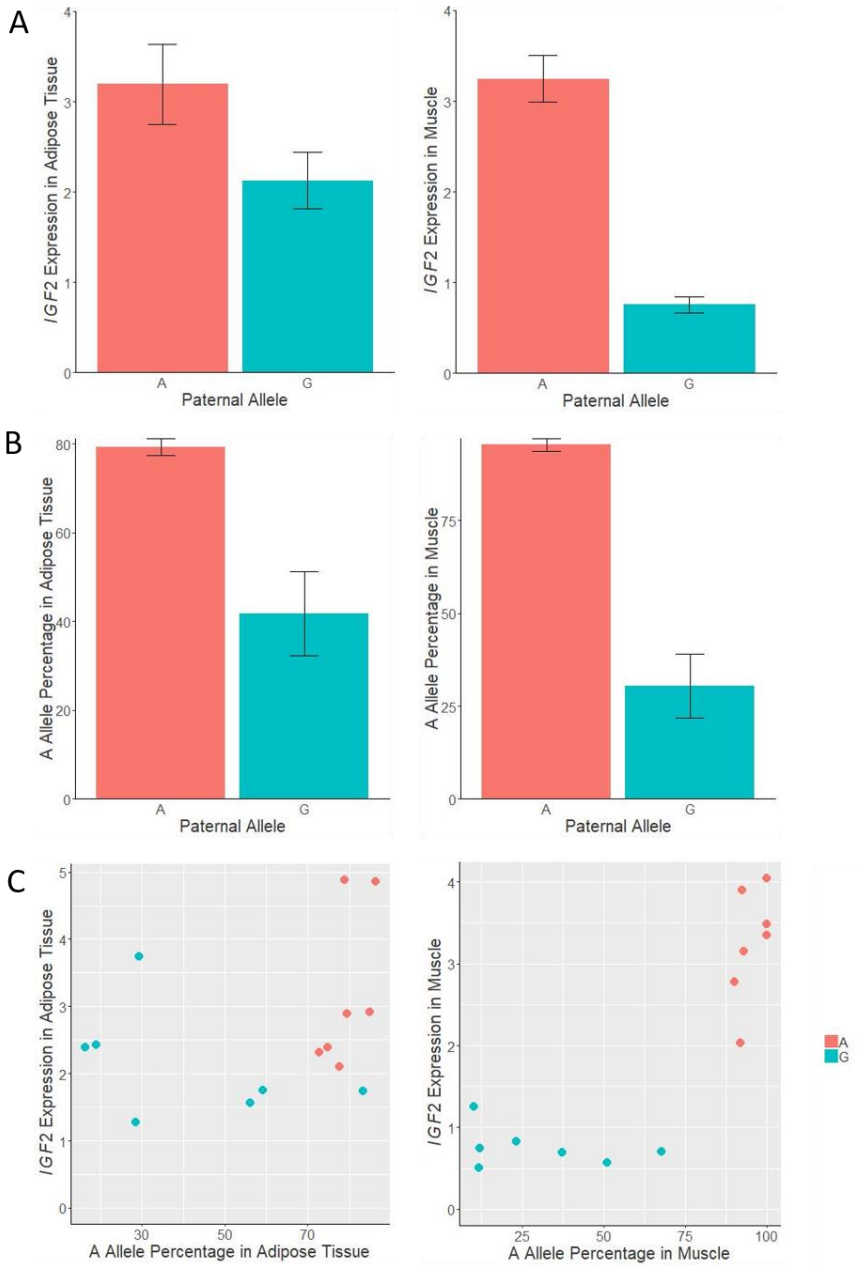


Fig 4. Plots of relative quantification of *IGF2* gene expression and allele percentage in muscle and adipose tissue according to the inherited paternal allele, and scatterplot combining *IGF2* gene expression and allele percentage in both tissues according to the paternal allele. The analysis was done in animals where paternally inherited allele was deduced. Data of *IGF2* gene expression represents means ± standard error of mean (SEM). Data for *A* allele are presented as percentage ± standard error.

However, these results may not agree with the adipose tissue *IGF2* gene expression comparison between the A^PG^M and A^MG^P genotypes (p-value=1.90x10^{-01}), in which no significant differences between these genotypes was observed, when the paternally

inherited allele was deduced from the genotypes of the parents in 355 backcrossed animals. Conversely, the animals with the *AA* genotype showed a higher *IGF2* gene expression in comparison with the other genotypes (Fig 5). Nonetheless, these results can be explained by a higher expression of the G allele in adipose tissue, in comparison with muscle, which may be in turn produced by a reduction of binding of the ZBED6 repressor.

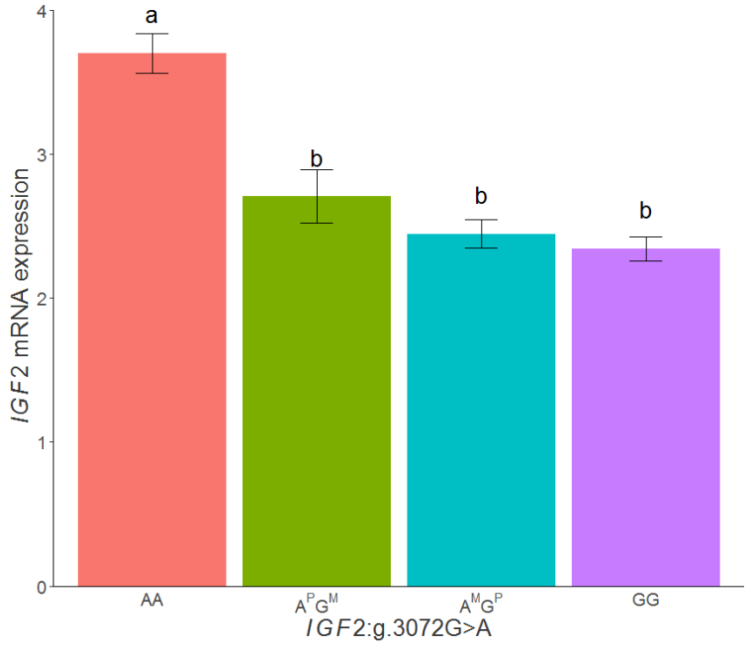


Fig 5. Plot of relative quantification of *IGF2* mRNA levels in adipose tissue according to the genotype of the *IGF2*:*g.3072G>A* polymorphism. The analysis was done in animals where paternally inherited allele was deduced, A^PG^M means a paternally inherited *A* allele and maternal inherited *G* allele, on the contrary, A^MG^P represents a maternal inherited *A* allele and paternal inherited *G* allele. Data represents means ± standard error of mean (SEM).

The structure of the pig *IGF2* gene consists in 10 exons but the mature form only contains the last three. The other exons, along with the four promoters included in the gene, are involved in the *IGF2* expression in a tissue specific manner [92]. For example, *IGF2* promoter 1 is used in liver instead of promoters 2, 3 and 4 that are used in muscle, being promoters tissue-dependent. Epigenetic regulation mechanisms, like imprinting status and its reflection in DNA methylation patterns are also completely different in each tissue [22], although no studies have been done in adipose tissue

either in humans and pigs [93]. It has been reported that *IGF2* is expressed from both parental alleles in liver, whereas imprinting has been described in mesodermal tissues such as skeletal muscle and kidney in foetal and adult animals [94]. Thus, we could assume that all the tissues coming from the mesoderm, including the adipose tissue, should present the same imprinting pattern. Supporting this hypothesis our results showed imprinting of the *IGF2* gene in muscle and adipose tissues. However, further studies are required to deepen the mechanism of regulation of *IGF2* gene expression in adipose tissue, which seems to play an important role in this tissue.

Sex and breed effects on *IGF2* gene expression

In order to identify if *IGF2* expression presents sexual dimorphism, the *IGF2* mRNA levels measured in adipose tissue of the 3BCs animals were analyzed according to sex. The obtained results showed that gene expression was higher in males (mean=3.12, SD=1.41, n=169) than in females (mean=2.71, SD=1.26, n=185), with significant differences (p-value=$1.19x10^{-4}$) and genotypic frequencies were balanced in the two sexes (Table 2). It is reported that some imprinted genes are related with sexual dimorphism in mice, including *IGF2*, in which gene expression is also higher in males than females and this can led to differences in body size between sexes [95].

Concerning the backcross effect, the highest *IGF2* gene expression was observed in BC1_DU (mean=3.66, SD=1.68) followed by BC1_PI (mean=2.64, SD=1.08) and BC1_LD (mean=2.36, SD=0.71). Significant differences were found between BC1_DU and BC1_LD (p-value=$1.66x10^{-3}$), and when comparing BC1_DU and BC1_PI (p-value=$1.55x10^{-5}$). On the contrary, no significant differences were obtained when gene expression of BC1_PI and BC1_LD was compared. These results are in accordance with the study of Redjuch *et al.* (2010), in which animals from Duroc, Large White, and Landrace breeds carrying the paternally derived *A* allele presented differences in *IGF2* gene expression, being higher in Duroc [96]. Hence, the differential *IGF2* gene expression among backcrosses may be explained by differences in genotypic frequencies (Table 2). Animals carrying the paternally derived *A* allele, that was deduced from the genotypes of the parents, were also analyzed according to the breed effect. The same results were obtained: highest *IGF2* gene expression corresponded to BC1_DU (mean=4.30, SD=1.77), followed by BC1_PI (mean=2.95, SD=1.05) and BC1_LD

(mean=2.55, SD=0.68) and significant differences were observed between the same breeds than in the previous study.

Association study for adipose tissue fatty acid composition and SSC2 polymorphisms

Since the identification of *IGF2*:*g.3072G>A* substitution as the causal mutation of the imprinted QTL for muscle growth, fat deposition and heart size [22] several association studies between the polymorphism and growth traits have been performed in different populations [27,84,85]. However, the association with FA composition in adipose tissue of the *IGF2:g.3072G>A* polymorphism has not been tested. In the present work, association analyses were carried out among 2,431 SNPs of SSC2, including the *IGF2:g.3072G>A* polymorphism, and FA composition measured in adipose tissue. The *IGF2:g.3072G>A* polymorphism was the most significantly associated with linoleic (C18:2(n-6); p-value=6.44x10^{-09}), hexadecanoic (C16:1(n-9); p-value=4.04x10^{-07}), oleic (C18:1(n-9); p-value=4.18x10^{-07}), α-linoleic (C18:3(n-3); p-value=3.30x10^{-06}), arachidonic (C20:4(n-6); p-value=9.82x10^{-08}) FAs and the MUFA/PUFA ratio (p-value=2.51x10^{-9}) (Fig 6).

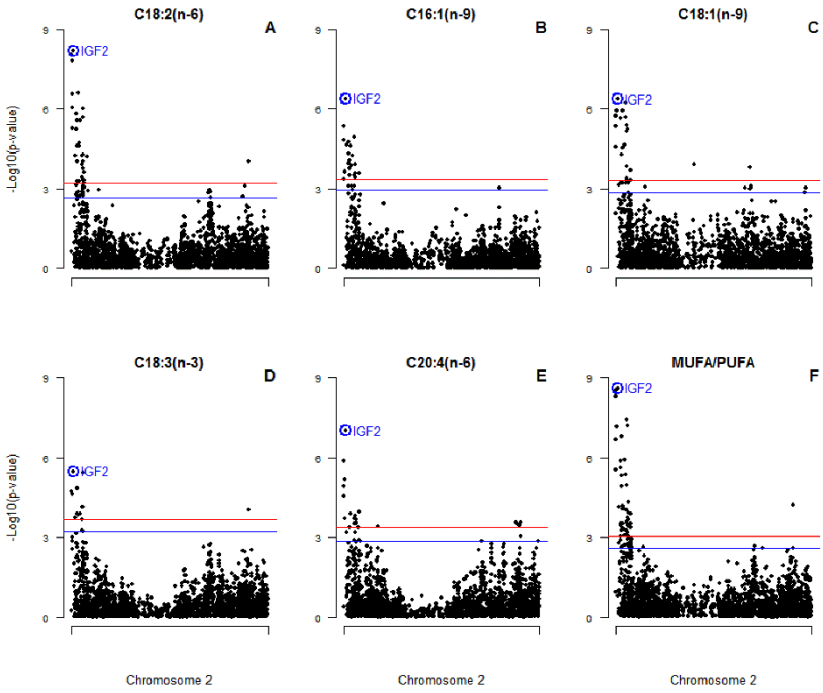


Fig 6. Plot of SSC2 SNPs association for significant FAs. (A) linoleic acid, (B) hexadecanoic acid, (C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb

based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

Correlations were performed between *IGF2* expression and FA composition measured in BF to deepen the relationship between gene expression and phenotypes. Our results showed a low correlation between the *IGF2* gene expression and FA composition. In general, a positive correlation between *IGF2* gene expression and the proportion of essential FAs, such as linoleic (r=0.21, p-value=$4.92x10^{-04}$) and α-linoleic (r=0.19, p-value=$3.56x10^{-04}$) FAs, in adipose tissue was observed. Conversely, a negative correlation with oleic FA (r=-0.21, p-value=$5.78x10^{-05}$) was shown. The imprinting model was also tested for FA composition and we neither could see FA content significant difference in the heterozygous genotype depending on which allele comes from the father.

SSC2 association studies were also performed independently in each backcross. In BC1_LD the *IGF2:g.3072G>A* polymorphism was not significantly associated with FA composition, and this could be also explained by the differences in the allele frequency of the SNP explained before. However, other significant polymorphisms of SSC2 were identified for linoleic (*rs81355859*, p-value=$3.22x10^{-07}$), hexadecanoic (*rs81322199*, p-value=$9.63x10^{-07}$), oleic (*rs81287787*, p-value=$4.39x10^{-06}$) and α-linoleic (*rs81316644*, p-value=$8.04x10^{-06}$) acids as well as for the MUFA/PUFA ratio (*rs81355859*, p-value=$1.02x10^{-06}$) (S3 Fig). In BC1_DU, there were not significant polymorphisms associated with the FA composition measured in adipose tissue (S4 Fig). Finally, in BC1_PI the *IGF2:g.3072G>A* polymorphism was the most significant associated SNP with linoleic (C18:2(n-6); p-value=$1.79x10^{-05}$), oleic (C18:1(n-9); p-value=$1.04x10^{-07}$), and arachidonic (C20:4(n-6); p-value=$2.79x10^{-05}$) FAs and the MUFA/PUFA ratio (p-value=$1.67x10^{-7}$) (S5 Fig), while the SNP *rs81312355* was the most significant associated polymorphism for the α-linoleic FA (p-value=$3.80x10^{-05}$).

The comparison among the 3BCs and the backcross specific studies showed that the *IGF2:g.3072G>A* polymorphism was not always the most significant SSC2 SNP. In BC1_LD animals other SNPs were more significant, in BC1_DU animals no significant SNPs were identified, and in BC1_PI animals *IGF2:g.3072G>A* was only the most significant SNP in four of the six traits described in the 3BCs animals. These results suggest that other variants associated with adipose tissue FA composition are segregating in specific backcrosses, mainly in BC1_LD.

Little is known about the relationship between the *IGF2* gene and lipid metabolism but it has been described that the IGF2 mRNA binding protein p62/IGF2BP2-2 is related with FA elongation in human liver disease [97]. Adipose tissue, which is the principal organ involved in the FA synthesis [98], has a high FA content, specifically of PUFA such as linoleic and α-linoleic FAs. These essential FAs are only provided by the diet and are readily stored in adipose tissue [99]. Besides, it was reported that there is an inverse relationship between the amount of α-linoleic in BF and the BF thickness, and this trait is related to fat quality in terms of firmness and the degree of cohesiveness within lean and fat tissues [1]. On the other hand, oleic acid is the most abundant MUFA in pork, comprising nearly 35%-45% of total FAs content. It is associated with consumer's acceptability of high quality cured products, in terms of organoleptic, technological and nutritional values of meat [100]. To study the effect of *IGF2:g.3072G>A* polymorphism in BF thickness, an association study was performed in 330 3BCs animals. No significant associations were found between the *IGF2:g.3072G>A* polymorphism and the fat measure (S6 Fig). The most significant SNP on SSC2 was *rs81214179*, which is located at 8.9 Mb (p-value=$1.57x10^{-06}$), where three desaturases involved in the synthesis of highly unsaturated FAs from essential FAs provided by the diet [99] were mapped: *fatty acid desaturase 1* (*FADS1*), *fatty acid desaturase 2* (*FADS2*) and *fatty acid desaturase 3* (*FADS3*).

In summary, according to the *IGF2*:*g.3072G>A* polymorphism, homozygous *AA* animals presented the highest *IGF2* gene expression in adipose tissue, a higher percentage of PUFA and a lower MUFA content in comparison to the other two genotypes. The association of the *IGF2:g.3072G>A* polymorphism with some relevant FAs suggest that IGF2 plays a role in the variability of FA composition in adipose tissue. Nevertheless,

we cannot exclude that other proximal genetic variants, such as polymorphisms located in the desaturases genes, are affecting FA content. Hence, further works are required to deepen the study of this complex SSC2 region. Considering that some studies in human reported the involvement of IGF2 in the metabolism and body fat composition, this gene could also play a physiological role in pig adipose tissue.

Conclusions

In the present work, the *AA* genotype *IGF2*:*g.3072G>A* polymorphism has been associated with a higher *IGF2* gene expression in BF adipose tissue. In addition, the *IGF2* gene expression in adipose tissue is explained by an imprinting model. Finally, the polymorphism was significantly associated with FA composition measured in BF and animals carrying the *A* allele showed a higher PUFA and lower MUFA content, although there may be other genetic variants affecting FA content. Hence, *IGF2* gene can play a relevant role in pig adipose tissue.

Supporting information

S1 Table. Primers used for IGF2 gene expression quantification by RT-qPCR.

S1 Fig. GWAS plot of adipose tissue *IGF2* gene expression in BC1_LD. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

S2 Fig. GWAS plot of adipose tissue *IGF2* gene expression in BC1_DU using *IGF2*:*g.3072G>A* polymorphism as a fixed effect. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value <0.1).

S3 Fig. Plot of SSC2 SNPs association for significant FAs in BC1_LD. (A) linoleic acid, (B) hexadecanoic acid, (C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F)

MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The IGF2:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

S4 Fig. Plot of SSC2 SNPs association for significant FAs in the BC1_DU. (A) linoleic acid, (B) hexadecanoic acid, (C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

S5 Fig. Plot of SSC2 SNPs association for significant FAs in the BC1_PI. (A) linoleic acid, (B) hexadecanoic acid,(C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2:g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.

S6 Fig. GWAS plot of BF thickness measure in the 3BCs animals. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value <0.1).

Acknowledgments

We wish to thank all of the members of the INIA, IRTA, and UAB institutions who contributed to the generation of the animal material used in this work. We are grateful to M. Costa for her contribution in the pyrosequencing analysis.

Author Contributions

Conceptualization: Lourdes Criado-Mesas, Maria Ballester, Josep M. Folch.

Formal analysis: Lourdes Criado-Mesas, Maria Ballester, Josep M. Folch.

Funding acquisition: Ana Isabel Fernández, Josep M. Folch.

Investigation: Lourdes Criado-Mesas, Maria Ballester, Daniel Crespo-Piazuelo, Anna Castello´, Rita Benítez, Josep M. Folch.

Methodology: Lourdes Criado-Mesas, Maria Ballester, Josep M. Folch. Resources: Ana Isabel Fernández, Josep M. Folch.

Supervision: Maria Ballester, Josep M. Folch.

Visualization: Lourdes Criado-Mesas.

Writing – original draft: Lourdes Criado-Mesas, Maria Ballester, Josep M. Folch. Writing – review & editing: Lourdes Criado-Mesas, Maria Ballester, Josep M. Folch.

Funding

This work was supported by the Spanish *Ministerio de Economía y Competitividad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) with project references: AGL2014-56369-C2 and AGL2017-82641-R. L. Criado-Mesas was financially supported by a FPI grant from the AGL2014-56369-C2 project. M. Ballester was funded with a "Ramón y Cajal" contract (RYC-2013-12573) from the Spanish Ministerio de Economía y Competitividad. D. Crespo-Piazuelo was funded by a "Formació i Contractació de Personal Investigador Novell" (FI-DGR) Ph.D grant from the Generalitat de Catalunya (ECO/1788/2014). We acknowledge the support of the Spanish *Ministerio de Economía y Competitividad* for the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-2015-0533) to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

References

1. Wood JD, Richardson RI, Nute GR, Fisher A V., Campo MM, Kasapidou E, et al. Effects of fatty acids on meat quality: A review. Meat Sci. 2004;66: 21–32. doi:10.1016/S0309-1740(03)00022-6

2. Wood JD, Enser M, Fisher A V., Nute GR, Sheard PR, Richardson RI, et al. Fat deposition, fatty acid composition and meat quality: A review. Meat Sci. 2008;78: 343–358. doi:10.1016/j.meatsci.2007.07.019

3. Jeon JT, Carlborg O, Törnsten a, Giuffra E, Amarger V, Chardon P, et al. A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. Nat Genet. 1999;21: 157–158. doi:10.1038/5938

4. Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R, et al. An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. Nat Genet. 1999;21: 155–156. doi:10.1038/5935

5. Smith EP. Insulin-Like Growth Factors and Skeletal Growth : 2015;84: 4349–4354.

6. Pavelic K, Buković D, Pavelić J. The role of insulin-like growth factor 2 and its receptors in human tumors. Mol Med. 2002;8: 771–80.

7. Livingstone C, Borai A. Insulin-like growth factor-II: Its role in metabolic and endocrine disease. Clin Endocrinol (Oxf). 2014;80: 773–781. doi:10.1111/cen.12446

8. Cianfarani S. Insulin-like growth factor-II: New roles for an old actor. Front Endocrinol (Lausanne). 2012;3: 1–4. doi:10.3389/fendo.2012.00118

9. Alfares MN, Perks CM, Hamilton-Shield JP, Holly JMP. Insulin-like growth factor-II in adipocyte regulation: depot-specific actions suggest a potential role limiting excess visceral adiposity. Am J Physiol Metab. 2018;315: E1098–E1107. doi:10.1152/ajpendo.00409.2017

10. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al.

A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature. 2003;425: 832–836. doi:10.1038/nature02064

11. Markljung E, Jiang L, Jaffe JD, Mikkelsen TS, Wallerman O, Larhammar M, et al. ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. PLoS Biol. 2009;7. doi:10.1371/journal.pbio.1000256

12. Gardan D, Gondret F, Van den Maagdenberg K, Buys N, De Smet S, Louveau I. Lipid metabolism and cellular features of skeletal muscle and subcutaneous adipose tissue in pigs differing in IGF-II genotype. Domest Anim Endocrinol. 2008;34: 45–53. doi:10.1016/j.domaniend.2006.10.001

13. Jungerius BJ, van Laere A-S, Te Pas MFW, van Oost B a, Andersson L, Groenen M a M. The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan x European white pig intercross. Genet Res. 2004;84: 95–101. doi:10.1017/S0016672304007098

14. Estellé J, Mercadé A, Noguera JL, Pérez-Enciso M, Óvilo C, Sánchez A, et al. Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross. J Anim Sci. 2005;83: 2723–2728. doi:83:2723-2728

15. Aslan O, Hamill RM, Davey G, McBryan J, Mullen AM, Gispert M, et al. Variation in the IGF2 gene promoter region is associated with intramuscular fat content in porcine skeletal muscle. Mol Biol Rep. 2012;39: 4101–4110. doi:10.1007/s11033-011-1192-5

16. López-Buesa P, Burgos C, Galve A, Varona L. Joint analysis of additive , dominant and first-order epistatic effects of four genes (IGF2 , MC4R , PRKAG3 and LEPR) with known effects on fat content and fat distribution in pigs. 2013; 133–137. doi:10.1111/age.12091

17. Sambrook, J., Fritsch, E. E. & Maniatis T. In Molecular Cloning: A Laboratory Manual 2nd edn. [Internet].

18. Pérez-Enciso M, Clop a, Noguera JL, Ovilo C, Coll a, Folch JM, et al. A QTL on pig chromosome 4 affects fatty acid metabolism : evidence from an Iberian by Landrace intercross The online version of this article , along with updated information and services , is located on the World Wide Web at : A QTL on pig chromosome 4 af. J Anim Sci. 2000;78: 2525–2531.

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

20. Puig-Oliveras A, Revilla M, Castelló A, Fernández AI, Folch JM, Ballester M. Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. Sci Rep. Nature Publishing Group; 2016;6: 31803. doi:10.1038/srep31803

21. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001;25: 402–408. doi:10.1006/meth.2001.1262

22. Ramayo-Caldas Y, Ballester M, Fortes MR, Esteve-Codina A, Castelló A, Noguera JL, et al. From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. BMC Genomics. 2014;15: 232. doi:10.1186/1471-2164-15-232

23. Ballester M, Cordón R, Folch JM. DAG expression: High-throughput gene expression analysis of real-time PCR data using standard curves for relative quantification. PLoS One. 2013;8: 8–12. doi:10.1371/journal.pone.0080385

24. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. J Comput Graph Stat. 1996;5: 299–314. doi:10.1080/10618600.1996.10474713

25. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. Nature Publishing Group; 2012;44: 821–4. doi:10.1038/ng.2310

26. Benjamini and Hochberg. Controlling the False Discovery Rate : A Practical and

Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 Published by : 1995;57: 289–300.

27. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics. 2010;26: 2069–2070. doi:10.1093/bioinformatics/btq330

28. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: An innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015;43: W589–W598. doi:10.1093/nar/gkv350

29. Braunschweig MH, Van Laere AS, Buys N, Andersson L, Andersson G. IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3. Genomics. 2004;84: 1021–1029. doi:10.1016/j.ygeno.2004.09.006

30. Chen M, Manley JL. Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. Nat Rev Mol Cell Biol. Nature Publishing Group; 2009;10: 741–754. doi:10.1038/nrm2777

31. Chao W, D'Amore PA. IGF2: Epigenetic regulation and role in development and disease. Cytokine Growth Factor Rev. 2008;19: 111–120. doi:10.1016/j.cytogfr.2008.01.005

32. Ruan W, Lai M. Insulin-like growth factor binding protein: A possible marker for the metabolic syndrome? Acta Diabetol. 2010;47: 5–14. doi:10.1007/s00592-009-0142-3

33. Li C, Bin Y, Curchoe C, Yang L, Feng D, Jiang Q, et al. Genetic imprinting of H19 and IGF2 in domestic pigs (Sus scrofa). Anim Biotechnol. 2008;19: 22–27. doi:10.1080/10495390701758563

34. Martínez JA, Milagro FI, Claycombe KJ, Schalinske KL. Epigenetics in Adipose Tissue , Obesity , Weight Loss , and Diabetes 1 , 2. Adv Nutr. 2014;5: 71–81.

doi:10.3945/an.113.004705.71

35. Braunschweig MH, Owczarek-Lipska M, Stahlberger-Saitbekova N. Relationship of porcine IGF2 imprinting status to DNA methylation at the H19 DMD and the IGF2 DMRs 1 and 2. BMC Genet. BioMed Central Ltd; 2011;12: 47. doi:10.1186/1471-2156-12-47

36. Faisal M, Kim H, Kim J. Sexual differences of imprinted genes' expression levels. Gene. Elsevier B.V.; 2014;533: 434–438. doi:10.1016/j.gene.2013.10.006

37. B. R. Expression of IGF1 and IGF2 genes in muscles during development of pigs representing five different breeds. Society. 2005;11: 1405–1417.

38. Laggai S, Kessler SM, Boettcher S, Lebrun V, Gemperlein K, Lederer E, et al. The *IGF2* mRNA binding protein p62/IGF2BP2-2 induces fatty acid elongation as a critical feature of steatosis. J Lipid Res. 2014;55: 1087–1097. doi:10.1194/jlr.M045500

39. Jansen GR, Hutchon CF, Zanetti ME. Studies on lipogenesis in vivo. Effect of dietary fat or starvation on conversion of [14]glucose into fat ad turnover of newly synthsized fat. Biochem J. Portland Press Ltd; 1966;99: 323–32.

40. Nakamura MT, Nara TY. Structure, Function, and Dietary Regulation of Δ6, Δ5, and Δ9 Desaturases. Annu Rev Nutr. 2004;24: 345–376. doi:10.1146/annurev.nutr.24.121803.063211

41. Hong J, Kim D, Cho K, Sa S, Choi S, Kim Y, et al. Effects of genetic variants for the swine FABP3, HMGA1, MC4R, IGF2, and FABP4 genes on fatty acid composition. Meat Sci. Elsevier B.V.; 2015;110: 46–51. doi:10.1016/j.meatsci.2015.06.011

PAPER III

Expression analysis of porcine miR-33a/b in liver, adipose tissue and muscle and its role in fatty acid metabolism

Lourdes Criado-Mesas[1], *et al*.

[1]Departament de Genòmica Animal, Centre de Recerca en Agrigenòmica (CRAG), CSIC-IRTA-UAB-UB, Barcelona, Spain.

E-mail: lourdes.criado@cragenomica.es

Manuscript in preparation

Abstract

mir-33a and *mir-33b* are co-transcribed with the *SREBF2* and *SREBF1* transcription factors, respectively. The main role of *SREBF1* is the regulation of genes involved in fatty acid metabolism, while *SREBF2* regulates genes participating in cholesterol biosynthesis and uptake. Our objective was to study the expression of both miR-33a and miR-33b in liver, adipose tissue and muscle to better understand their role in lipid metabolism in pigs. We observed different tissue expression patterns for both miRNAs, suggesting different expression regulatory mechanisms according to tissue. In adipose tissue and muscle a high expression correlation between miR-33a and miR-33b was observed, suggesting a similar regulation and regulatory role, while a lower correlation in liver may indicate different functions for each miR-33 family member. The expression analysis of *in-silico* predicted target-lipid related genes showed negative correlations between miR-33b and *CPT1A* expression in liver. Conversely, positive correlations between miR-33a and *PPARGC1A* and *USF1* gene expression in liver were observed. These results are in accordance with the different function of the miR-33a and miR-33b described in liver, pointing to a decrease of lipolysis pathways and a consequent activation of cholesterol and an increase of lipogenesis pathways. Finally, positive and negative correlations between miR-33a/b and saturated fatty acid (SFA) and polyunsaturated fatty acid (PUFA) content, respectively, suggested a role of these genes in the fatty acid composition of the adipose tissue.

Introduction

Pork is one of the most consumed meats in the world, being meat quality a relevant trait for both the meat industry and consumers. Among meat quality characteristics, intramuscular fat (IMF) content and fatty acid (FA) composition determine not only meat flavour, tenderness, firmness and juiciness, but also the healthiness of the product (Wood and Whittemore, 2007; Wood *et al.*, 2008). In addition, pig is considered a good animal model for biomedical research because of its similarities with humans, and has been used to identify drug targets against human diseases, such as obesity (Rocha and Plastow, 2006).

Liver, adipose tissue and skeletal muscle are the principal metabolic organs involved in the regulation of lipid metabolism and, therefore, play an important role in the determination of IMF content and FA composition. In pigs, liver is participating in the synthesis and secretion of very low-density proteins, de novo cholesterol synthesis and fatty acid β-oxidation. In addition, liver and adipose tissue are involved in de novo fatty acid synthesis (Nguyen *et al.*, 2008), with a higher contribution from adipose tissue. Moreover, adipose tissue is an organ acting in lipid storage and maintenance of metabolic homeostasis, and it is the major source of circulating free FAs (O'Hea and Leveille, 1969; Kershaw and Flier, 2004). Finally, muscle is a key place for glucose uptake and storage, and a reservoir of amino acids necessary for protein synthesis or energy production (Meyer *et al.*, 2002). The lipid metabolism pathways are cross-regulated among liver, adipose tissue and muscle, and have been extensively studied. In previous studies of our group, candidate genes involved in different metabolic pathways in liver, adipose tissue and muscle and affecting IMF content and FA composition in pigs were identified by using GWAS, RNA-Seq and co-association network approaches (Corominas *et al.*, 2012, 2013, 2015; Ramayo-Caldas *et al.*, 2012a; b, 2014; Puig-Oliveras *et al.*, 2014b; a). In addition, the expression and polymorphisms of several lipid-related candidate genes were studied (Estellé *et al.*, 2009; Corominas *et al.*, 2012, 2015; Ballester *et al.*, 2016; Puig-Oliveras *et al.*, 2016; Ballester, Puig-Oliveras, *et al.*, 2017; Ballester, Ramayo-Caldas, *et al.*, 2017; Revilla *et al.*, 2018; Criado-Mesas *et al.*, 2019).

Besides the transcriptional gene expression regulation, miRNAs have emerged as important post-transcriptional regulators of genes involved in lipid metabolism in different porcine tissues (Song *et al.*, 2018). microRNAS (miRNAs) are small RNA molecules that prevent the production of proteins or degrade the mRNA (Reddy *et al.*, 2009). They play important roles in diverse regulatory pathways of many cellular processes and diseases. Members of the miR-33 family, which includes *mir-33a* and *mir-33b* are located in *SREBF2* intron 13 and *SREBF1* intron 16, respectively, and were reported to be co-transcribed with their host genes. SREBP transcription factors are well-known master regulators of lipid homeostasis. *SREBF1* regulates genes mainly involved in fatty acid metabolism, while *SREBF2* regulates genes involved in cholesterol biosynthesis and uptake (Shimano, 2001; Dávalos *et al.*, 2011). Pig miR-33a/b sequences differ only in three nucleotides, have the same seed sequence, and are conserved with the human homologous genes. In line with the regulatory functions of its host genes, human miR-33b was reported to regulate the insulin signalling pathway and glucose synthesis, which affected gluconeogenesis pathways (Ramirez *et al.*, 2013; Zhang *et al.*, 2019), and miR-33a was involved in the regulation of genes of cholesterol synthesis (Horie *et al.*, 2010; Rayner *et al.*, 2010). In pigs, only miR-33b has been reported to play an important role in adipogenesis and lipogenesis in adipose tissue (Taniguchi *et al.*, 2014).

The aim of this work was to study the expression of miR-33a and miR-33b in the three main metabolic tissues, liver, adipose tissue and muscle, and their effect on FA composition measured in muscle and adipose tissue, to better understand their role in lipid metabolism in swine.

Material and methods

Animal samples

The animal material used in this study comes from the IBMAP experimental cross population, which was generated by crossing three Iberian (Guadyerbas line) boars with 31 Landrace sows and after that five F_1 males were backcrossed with 25 Landrace

sows (BC1_muscle) (Pérez-Enciso *et al.*, 2000). Here, we analysed 42 pigs from the BC1_LD (25% Iberian x 75% Landrace) generation. All animals were fed *ad libitum* with a cereal-based commercial diet and maintained under intensive conditions. After slaughter, liver, adipose tissue and *Longissimus dorsi* muscle samples were collected and immediately snap-frozen in liquid nitrogen and stored at -80°C until analysis. Animal care and procedures were performed following national and institutional guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA- Institut de Recerca i Tecnologia Agroalimentàries).

Phenotypic data

Composition of FAs with 12-22 carbons was determined in muscle (Ramayo-Caldas *et al.*, 2012b) and backfat adipose tissue (Muñoz *et al.*, 2013) using a protocol based on gas chromatography of methyl esters (Pérez-Enciso *et al.*, 2000). Afterwards, the percentage of the content of each FA was calculated in addition to the overall percentage of saturated FAs (SFA), monounsaturated FAs (MUFA) and polyunsaturated FAs (PUFA).

Reverse transcription quantitative PCR (RT-qPCR)

Total RNA was purified from 50 mg of liver tissue and from 150 mg of adipose tissue samples and was directly homogenized in 1 mL of TRIzol Reagent with a polytron device. In the case of muscle (*Longissimus dorsi*) samples, 100 mg were submerged in liquid nitrogen and ground with a mortar and a pestle before adding 1 mL of TRIzol. Next, 200 µl of chloroform were added and samples were centrifuged to separate the nucleic acids and proteins from the RNA. Supernatant was collected to a new tube and total RNA was precipitated by adding 500 µl of isopropanol and washed with 75% ethanol (Chomczynski and Sacchi, 1987). Finally, the RNA was resuspended with 100 µl in liver samples and 50 µl in adipose tissue and muscle samples with RNAse free water. RNA concentration and purity was measured using a NanoDrop ND-1000 spectrophotometer (NanoDrop products) and RNA integrity was checked by using an Agilent Bioanalyzer-2100 (Agilent Technologies).

Total RNA was reverse transcribed into cDNA with the Taqman Advanced miRNA cDNA synthesis kit (Applied Biosystems) by using 2 µl (5 ng/ µl) of total RNA in a final reaction volume of 30 µl. Then, 5 µl of the resulting RT reactions were amplified in a final volume of 50 µl following manufacturer's instructions. Finally, cDNA was diluted 1/10 for RT-qPCR. A negative control was made for each tissue with no reverse transcriptase added. cDNA was stored at -20ºC until use.

Pre-designed Taqman MicroRNA Assays (Applied Biosystems) were used for hsa-miR-33a, hsa-miR-33b, hsa-miR-let7a and hsa-miR-26a. Primers were designed for *SREBF2* gene and reported in Supplementary Table 1. Relative quantification of hsa-miR-33a, hsa-miR-33b and *SREBF2* by real-time PCR (qPCR) was performed in a QuantStudioTM 12K Flex Real-Time PCR System (ThermoFisher Scientific) using a 384-well plate and all reactions were done per triplicate. miR-let7a and miR-26a were used as porcine reference miRNAs and were chosen according to the bibliography (Timoneda *et al.*, 2012), and *ACTB* and *TBP* were used as porcine reference mRNAs (mRNA primers were reported in Supplementary Table 1) (Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017b; Revilla *et al.*, 2018). The PCR thermal cycle was: 2 min at 50ºC, 10 min at 95 °C, 40 cycles of 15 sec at 95 °C and 1 min at 60 °C. Moreover, a melting profile (95 ºC for 15 sec, 60 ºC for 15 sec and a gradual increase in temperature with a ramp rate of 1% up to 95 ºC) was added following the thermal cycling protocol, to assess for the specificity of the reactions. Data was analysed with the ThermoFisher Cloud software 1.0 (Applied Biosystems) and the $2^{-\Delta Ct}$ (Livak and Schmittgen, 2001) method was applied. *SREBF1* mRNA expression data in liver, adipose tissue, and muscle was previously generated by Ballester *et al.* 2017, Revilla *et al.* 2018 and Puig-Oliveras *et al.* 2016, respectively.

Statistical analysis

Normalization of data was checked using Shapiro-Wilk test in R (R Core Team, 2018) and $\log_2$ transformation of the NQ value was applied if necessary. Means were compared using Tukey Honest Significant Difference (HSD) test (Dubitzky *et al.*, 2013). Pearson's correlations were performed among target gene expression and miR-33b quantification using R software. Porcine mRNA 3'UTRs sequences were downloaded from the Ensembl database and Seqkit tool (Shen *et al.*, 2016) was used to search by

homology those mRNA 3'UTR sequences matching with 7mer seed miRNA sequence. Additionally, we assessed the conservation and confidence of the miR-33a/b putative target sites among other mammal species by using the TargetScan webserver (Agarwal *et al.*, 2015).

Results

miR-33a and miR-33b expression in liver, adipose tissue and muscle

In the current study, miR-33a and miR-33b expression quantification was performed in liver, adipose tissue and muscle of 42 pigs (Figure 1).

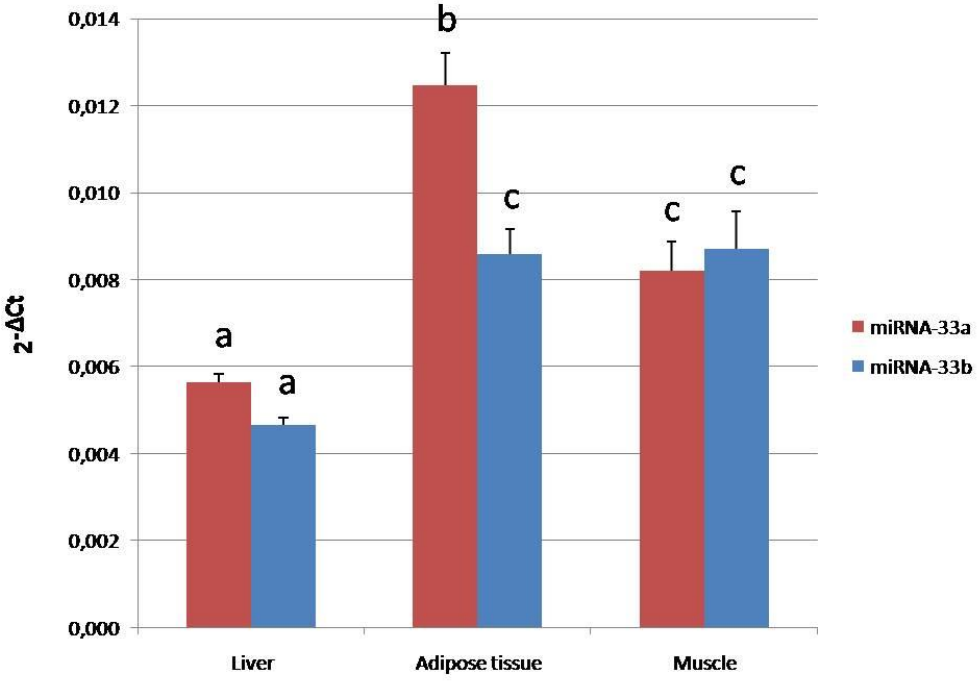


Figure 1. miR-33a and miR-33b expression in liver, adipose tissue and muscle. Data represents $2^{-\Delta Ct}$ mean ± standard error of the mean (SEM). Superscript letters represent significant differences obtained after a Tukey's HSD test.

The highest level of miR-33a expression was observed in adipose tissue, followed by muscle and liver. By contrast, miR-33b showed a higher expression in muscle and

adipose tissue in comparison to liver. Between the two miR-33 genes, miR-33a presented a higher expression than miR-33b in adipose tissue (*p*-value=1.16x10^{-04}).

Correlations between miR-33a and miR-33b among tissues were calculated (Figure 2), showing a high correlation in muscle (r= 0.92, *p*-value = 2.76x10^{-16}) and adipose tissue (r=0.83, *p*-value = 9.60x10^{-11}). Conversely, a lower correlation between miR-33a and miR-33b was observed in liver (r=0.36, *p*-value = 2.25x10^{-02}). Furthermore, correlations among tissues were only significant for liver and adipose tissue miR-33b expressions (r=0.32, *p*-value=4.51x10^{-02}).

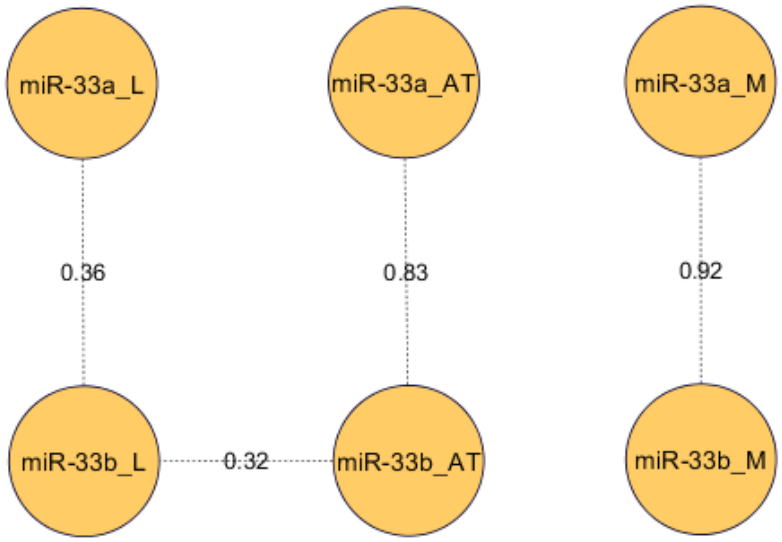


Figure 2. Pearson correlations between miR-33a and miR-33b-expression in liver (L), adipose tissue (AT) and *longissimus dorsi* muscle (M). Only significant correlations were represented.

miR-33a/b expression correlations with *SREBF2* and *SREBF1* respectively

It is well-known that both miR-33a and miR-33b are located in intronic regions of *SREBF2* and *SREBF1* genes, respectively. In order to study if both miR-33a and miR-33b are co-transcribed with their host genes, correlations among their expression levels in

the three tissues were calculated. While no significant correlations were found between miR-33b and *SREBF1* gene in any tissue, a significant positive correlation between the expression of miR-33a and *SREBF2* in liver was found (r=0.5, *p*-value=$1.12x10^{-03}$).

Association among the expression levels of miR-33 and target genes

Considering the relevant role that miR-33 members play in lipid and cholesterol metabolism, we wanted to study the association between expression levels of lipid-related genes and miR-33a and miR-33b. To this purpose, previously published mRNA expression data of 45 lipid-related genes in liver, adipose tissue and muscle was used (Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017b; Revilla *et al.*, 2018). In these works, gene expression was quantified by qPCR in a set of animals which included the 42 animals of the present work. Supplementary Table 2 summarizes the list of genes analysed in each tissue and shows that most of the genes match among tissues. To identify potential binding sites for miR-33 in the 84 lipid-related genes with expression data, their porcine 3'UTRs sequences were downloaded from the Ensembl database and searched for homology with the 7mer seed miR-33 sequence using the Seqkit tool (Shen *et al.*, 2016). Fifteen genes contained the 7mer seed miRNA sequence in their 3'UTR and are listed in Table 1.

Table 1. Genes with the 7mer seed miR-33 sequence in their 3'UTR and tissues in which expression data is available

Gene	Tissues
ACSM5	Liver, adipose tissue and muscle
ADIPOQ	Adipose tissue
CPT1A	Liver
CROT	Liver, adipose tissue and muscle
HNF4A	Liver
LIPC	Liver and adipose tissue
MGLL	Muscle and adipose tissue
MLXIPL	Liver, adipose tissue and muscle
NCOA1	Muscle
NR1H3	Liver and adipose tissue
PPARGC1A	Liver and muscle
PRKAA1	Muscle
SETD7	Muscle
SCAP	Adipose tissue
USF1	Liver and adipose tissue

Moreover, the 3'-UTR target sites conservation between human and pig was evaluated i*n silico* using the TargetScan algorithm. The *CPT1A*, *CROT*, *LIPC*, *NCOA1*, *PRKAA1*, and *SETD7* predicted miR-33 target sites were highly conserved among species and showed a context++ score higher than 70% percentile. This score is considered as confidently cross-validated and shows the probability of all the predicted target sites to be biologically functional (Agarwal *et al.*, 2015).

Low to moderate significant correlations were found among miR-33a and miR-33b and their target genes in the three tissues (Supplementary Table 3). It is relevant to highlight the negative correlation observed between miR-33b and *CPT1A* in liver although it not reaches statistical significance (*p*-value=0.086). Also, positive

correlations were observed between most of the genes and miR-33a/b, although statistically significant correlations were only obtained between miR-33a and both *PPARGC1A* and *USF1* expression values in liver (*p*-value < 0.05).

Association between miR-33a and miR-33b expression and fatty acid composition

The association among miR-33a and miR-33b measured in the three tissues and FA composition measured in backfat adipose tissue and muscle was studied by Pearson's correlation. While no significant correlations were found between miR-33a/b and FA composition measured in muscle, significant correlations were found between miR-33a/b measured in liver and adipose tissue and FA composition measured in adipose tissue (Table 2). Specifically, liver miR-33a expression was positively correlated with SFA total content, and negatively correlated with linoleic (C18:2(n-6)) and eicosatrienoic (C20:3(n-6)) fatty acids, as well as the PUFA total content in adipose tissue. In addition, liver miR-33b expression showed positive correlations with myristic (C14:0) and palmitic (C16:0) fatty acids, and a negative correlation with eicosatrienoic (C20:3(n-6)) fatty acid in adipose tissue. The expression of both miR-33a/b in adipose tissue was positively correlated with stearic (C18:0) fatty acid and SFA total content, while negative correlations were found with the PUFA total content, along with linoleic (C18:2(n-6)) fatty acid. Adipose tissue miR-33a expression was also negatively correlated with eicosatrienoic (C20:3(n-6)) fatty acid.

Table 2. Summary of correlation values for both miRNAs measured in liver and adipose tissue and FA composition measured in backfat adipose tissue. *P*-values are indicated in brackets and * means statistically significant (*p*-value < 0.05).

Liver		
FA	**miR-33a**	**miR-33b**
C140	0.18 (2.60E-01)	0.38 (1.60E-02)*
C160	0.25 (1.22E-01)	0.34 (3.70E-02)
SFA	0.36 (2.44E-02)*	0.20 (2.34E-01)
C182n6	-0.40 (1.15E-02)*	-0.30 (6.65E-02)
C203n6	-0.35 (2.66E-02)*	-0.33 (3.91E-02)*
PUFA	-0.38 (1.63E-02)*	-0.29 (6.94E-02)

Adipose tissue		
FA	**miR-33a**	**miR-33b**
C180	0.35 (2.82E-02)*	0.35 (3.12E-02)*
SFA	0.32 (4.41E-02)*	0.42 (7.93E-03)*
C182n6	-0.41 (9.90E-03)*	-0.49 (2.03E-03)*
C203n6	-0.34 (3.42E-02)*	-0.29 (8.22E-02)
PUFA	-0.40 (1.09E-02)*	-0.48 (2.16E-03)*

Discussion

Since the miR-33 family has a relevant role in the regulation of genes involved in lipid metabolism pathways, in the current work, the expression of miR-33a and miR-33b in liver, adipose tissue and muscle, and its correlation with both *in-silico* predicted target lipid-related genes and fatty acid composition traits were studied.

Several studies in human and mice have reported that miR-33a and miR-33b are co-transcribed with their host genes, *SREBF2* and *SREBF1,* respectively (Gerin *et al.*, 2010;

Marquart *et al.*, 2010; Najafi-Shoushtari *et al.*, 2010; Rayner *et al.*, 2010; Dávalos *et al.*, 2011). However, a low correlation between miR-33b and *SREBF1* gene expression has been reported in adipose tissue of pigs (Taniguchi *et al.*, 2014) and members of the miR-33 family are not co-regulated with their host genes in most tissues of chickens (Shao *et al.*, 2014). Accordingly, in the present study no significant correlation was observed between miR-33b and *SREBF1* gene expression in any tissue (Supplementary Table 4). On the contrary, miR-33a and *SREBF2* gene expression in liver showed a positive correlation (r=0.5; *p*-value=$1.12x10^{-03}$). Overall, these results suggest that both miRNAs are transcribed in a different way.

Additionally, the analysis of miR-33a and miR-33b expression in liver, adipose tissue and muscle revealed different expression patterns among tissues for both miRNAs. Similar results have been also reported in humans with different levels of miR-33a and miR33-b expression depending on tissue (Ludwig *et al.*, 2016), which suggest that different tissue-specific mechanisms are regulating the expression of miR-33a/b. Conversely, high correlations between miR-33a and miR-33b expression levels (r > 0.8) were obtained within the muscle and the adipose tissues, suggesting a similar regulation in the expression of both miRNAs in these tissues. In fact, taking into account that both miR-33a/b have the same seed sequence, we cannot discard that both miR-33a/b play a similar function in these tissues. However, different expression levels between miR-33a and miR-33b in adipose tissue were found. To the best of our knowledge, there are no published works regarding the role of miR-33a in the adipose tissue of pigs. A study published in humans determined that miR-33a was constitutively expressed while miR-33b expression increased during adipocyte differentiation (Price *et al.*, 2016). Contrarily, transfection of miR-33b in porcine subcutaneous preadipocytes downregulates adipose differentiation and lipid accumulation (Taniguchi *et al.*, 2014). Thus, further studies are necessary to better understand the role of miR33a in pig adipose tissue and determine if both miR-33a/b have different regulatory functions in this tissue.

A different expression pattern was observed for both miR-33a/b in liver, where the lowest expression levels and correlation values between miR-33a and miR-33b (r=0.36) were obtained. It has been reported that miR-33a and miR-33b work in collaboration

with their host genes regulating lipid metabolism in liver, and while miR-33a participates in the transcriptional control of genes involved in cholesterol pathways (Horie *et al.*, 2010; Marquart *et al.*, 2010; Najafi-Shoushtari *et al.*, 2010; Rayner *et al.*, 2010; Ramirez *et al.*, 2013), miR-33b was related with fatty acid oxidation and insulin signalling pathway (Gerin *et al.*, 2010; Dávalos *et al.*, 2011). In pigs, liver plays an important role in *de novo* cholesterol synthesis, lipogenesis and fatty acid oxidation (O'Hea and Leveille, 1969; Gondret, Ferré and Dugail, 2001; Kershaw and Flier, 2004; Nafikov and Beitz, 2007; Nguyen *et al.*, 2008).. In line with the low correlation values observed between both miR-33a/b in liver, miR-33b tended to be higher negatively correlated with *CPT1A* expression levels than miR-33a. Therefore, we could hypothesize that both miR-33a/b plays a different regulatory role in liver, with miR-33b being involved in FA β-oxidation. These is also supported by the positive correlations between miR-33a and *PPARGC1A* and *USF1* found in liver, because they are transcription factors involved in the regulation of several genes of fatty acid metabolism (Griffin and Sul, 2004; Lin, Handschin and Spiegelman, 2005; Finck and Kelly, 2006).

Finally, significant positive correlations between both miR-33a/b expressions in either liver and adipose tissue and SFAs and/or total SFA content, whereas negative correlations with PUFAs and/or total PUFA content were observed (Figure 3). This results were in accordance with other studies where an increase of PUFA content was related with a decrease of *SREBF* and consequently miR-33 and genes involved in lipogenesis in liver (Xu *et al.*, 1999; Clarke, 2001). Interestingly, previous studies of our group reported that BC1_LD animals with a higher content of PUFA measured in muscle increased the expression of genes involved in the fatty acid oxidation and cholesterol homeostasis and inhibits lipogenesis pathways in liver and adipose tissue (Ramayo-Caldas *et al.*, 2012a; Corominas *et al.*, 2013). These results are in accordance with the current ones because if the PUFA content increases, the expression of miR-33 and consequently genes involved in lipogenesis decrease while fatty acid oxidation increases.

On the other hand, it has been demonstrated that cholesterol increases the expression of lipogenic genes and the triglyceride content in mouse liver (Knight *et al.*, 2005),

among other metabolic changes such as liver lipid content and free fatty acid metabolism (Tsai, Romsos and Leveille, 1975). Hence, an interaction between cholesterol and lipogenesis pathways may explain the correlation between the miR-33a and the fatty acid composition measured in the adipose tissue.

Altogether, our results indicate the possible implication of miR-33 family in the determination of FA composition in adipose tissue. This is of great interest because SFAs are related to modern human diseases such as obesity, cancer and cardiovascular diseases, while PUFAs are more susceptible to be oxidized, which produces a reduction of meat quality (Webb and O'Neill, 2008; Wood *et al.*, 2008).

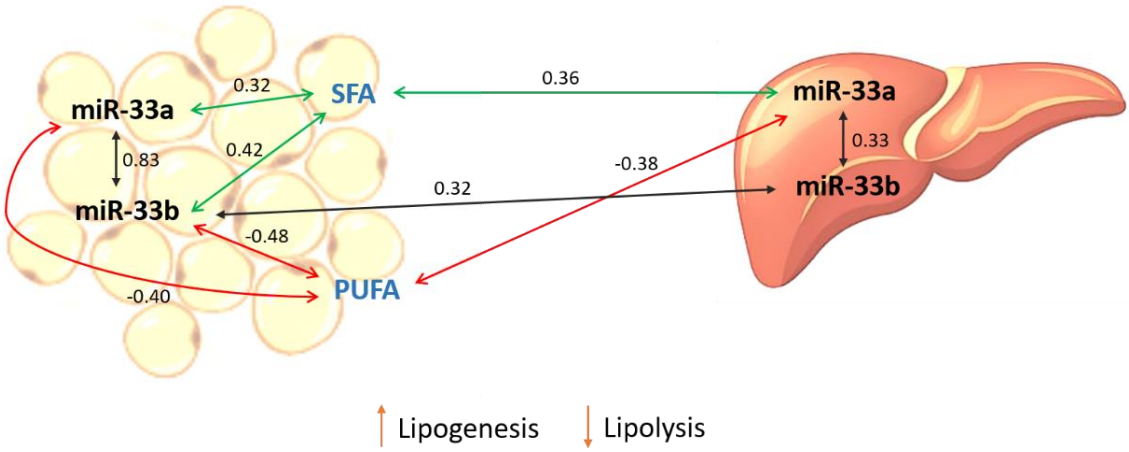


Figure 3. Schematic representation of correlation results between miR-33a/b expression measured in liver and adipose tissue and both target genes and FAs measured in adipose tissue.

Conclusions

In general, our results indicate that miR-33a and miR-33b are transcribed in a different manner and the miR-33a/b expression regulatory mechanisms are different according to tissue. miR-33a and miR-33b expression levels presented high correlations in adipose tissue and muscle which may indicate a similar regulation in these tissues. Conversely, a different expression pattern and low expression correlations between miR-33a and miR-33b in liver indicates different regulatory functions for both miRNAs

in this tissue. A negative correlation found between miR-33b and *CPT1A* expression in liver and positive correlations observed between miR-33a and *PPARGC1A* and *USF1* transcription factors reinforced the hypothesis that both miRNAs have different functions in liver and miR-33b is involved in FA-β-oxidation. In general, no significant correlations between the miR-33a/b and their target genes were found, so further studies are needed to determine the post-transcriptional miRNA regulation. Finally, the miR-33 family may be involved in the determination of FA composition in adipose tissue.

References

Agarwal, V., Bell, G. W., Nam, J. W. and Bartel, D. P. (2015) 'Predicting effective microRNA target sites in mammalian mRNAs', *eLife*, 4, pp. 1–38.

Ballester, M., Puig-Oliveras, A., Castelló, A., Revilla, M., Fernández, A. I. and Folch, J. M. (2017a) 'Association of genetic variants and expression levels of porcine FABP4 and FABP5 genes', *Animal Genetics*, 48(6), pp. 660–668.

Ballester, M., Ramayo-Caldas, Y., Revilla, M., Corominas, J., Castelló, A., Estellé, J., *et al.* (2017b) 'Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs', *Scientific Reports*, 7, p. 46539.

Ballester, M., Revilla, M., Puig-Oliveras, A., Marchesi, J. A. P., Castelló, A., Corominas, J., *et al.* (2016) 'Analysis of the porcine APOA2 gene expression in liver, polymorphism identification and association with fatty acid composition traits', *Animal genetics*, 47(5), pp. 552–559.

Chomczynski, P. and Sacchi, N. (1987) 'Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction', *Analytical Biochemistry*, 162(1), pp. 156–159.

Clarke, S. D. (2001) 'Nonalcoholic steatosis and steatohepatitis. I. Molecular mechanism for polyunsaturated fatty acid regulation of gene transcription', *Am J*

Physiol Gastrointest Liver Physiol, 281, pp. 865–869.

Corominas, J., Marchesi, J. A., Puig-Oliveras, A., Revilla, M., Estellé, J., Alves, E., *et al.* (2015) 'Epigenetic regulation of the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs', *Genetics Selection Evolution*, 47(20), pp. 1–11.

Corominas, J., Ramayo-Caldas, Y., Castelló, A., Muñoz, M., Ibáñez-Escriche, N., Folch, J. M., *et al.* (2012) 'Evaluation of the porcine ACSL4 gene as a candidate gene for meat quality traits in pigs', *Animal Genetics*, 43(6), pp. 714–720.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., *et al.* (2013) 'Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition', *BMC Genomics*, 14(843).

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Benítez, R., Fernández, A. I., *et al.* (2019) 'Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition', *PLOS ONE*, 14(8), p. e0220708.

Dávalos, A., Goedeke, L., Smibert, P., Ramírez, C. M., Warrier, N. P. and Andreo, U. (2011) 'miR-33a / b contribute to the regulation of fatty acid metabolism and insulin signaling', *Pnas*, 108(22), pp. 9232–9237.

Dubitzky, W., Wolkenhauer, O., Cho, K.-H. and Yokota, H. (eds) (2013) 'Tukey's HSD Test', in *Encyclopedia of Systems Biology*. p. 2303.

Estellé, J., Fernández, A. I., Pérez-Enciso, M., Fernández, A., Rodríguez, C., Sánchez, A., *et al.* (2009) 'A non-synonymous mutation in a conserved site of the MTTP gene is strongly associated with protein activity and fatty acid profile in pigs', *Animal Genetics*, 40(6), pp. 813–820.

Finck, B. N. and Kelly, D. P. (2006) 'PGC-1 coactivators : inducible regulators of energy metabolism in health and disease Find the latest version : Review series PGC-1 coactivators : inducible regulators of energy metabolism in health and disease', *Journal of Clinical Investigation*, 116(3), pp. 615–622.

Gerin, I., Clerbaux, L. A., Haumont, O., Lanthier, N., Das, A. K., Burant, C. F., *et al.* (2010)

'Expression of miR-33 from an SREBP2 intron inhibits cholesterol export and fatty acid oxidation', *Journal of Biological Chemistry*, 285(44), pp. 33652–33661.

Gondret, F., Ferré, P. and Dugail, I. (2001) 'ADD-1/SREBP-1 is a major determinant of tissue differential lipogenic capacity in mammalian and avian species', *Journal of Lipid Research*, 42(1), pp. 106–113.

Griffin, M. J. and Sul, H. S. (2004) 'Insulin regulation of fatty acid synthase gene transcription: Roles of USF and SREBP-1c', *IUBMB Life*, 56(10), pp. 595–600.

Horie, T., Ono, K., Horiguchi, M., Nishi, H., Nakamura, T., Nagao, K., *et al.* (2010) 'MicroRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (Srebp2) regulates HDL in vivo', *PNAS*, 107(40), pp. 17321–17326.

Kershaw, E. E. and Flier, J. S. (2004) 'Adipose tissue as an endocrine organ', *Journal of Clinical Endocrinology and Metabolism*, 89(6), pp. 2548–2556.

Knight, B. L., Hebbach, A., Hauton, D., Brown, A. M., Wiggins, D., Patel, D. D., *et al.* (2005) 'A role for PPARα in the control of SREBP activity and lipid synthesis in the liver', *Biochemical Journal*, 389(2), pp. 413–421.

Lin, J., Handschin, C. and Spiegelman, B. M. (2005) 'Metabolic control through the PGC-1 family of transcription coactivators', *Cell Metabolism*, 1(6), pp. 361–370.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method', *Methods*, 25(4), pp. 402–408.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., *et al.* (2016) 'Distribution of miRNA expression across human tissues', *Nucleic Acids Research*, 44(8), pp. 3865–3877.

Marquart, T. J., Allen, R. M., Ory, D. S. and Baldan, A. (2010) 'miR-33 links SREBP-2 induction to repression of sterol transporters', *Proceedings of the National Academy of Sciences*, 107(27), pp. 12228–12232.

Meyer, C., Dostou, J. M., Welle, S. L. and Gerich, J. E. (2002) 'Role of human liver, kidney, and skeletal muscle in postprandial glucose homeostasis', *American Journal of Physiology-Endocrinology and Metabolism*, 282(2), pp. E419–E427.

Muñoz, M., Rodríguez, M. C., Alves, E., Folch, J. M., Ibañez-Escriche, N., Silió, L., *et al.* (2013) 'Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data', *BMC Genomics*, 14(1).

Nafikov, R. A. and Beitz, D. C. (2007) 'Carbohydrate and lipid metabolism in the newborn.', *The Journal of Nutrition*, 137, pp. 702–705.

Najafi-Shoushtari, S., Kristo, F., Li, Y., Shioda, T., Cohen, D., Gerszten, R., *et al.* (2010) 'MicroRNA-33 and the SREBP Host Genes Cooperate to Control Cholesterol Homeostasis', *Science*, 328(405), pp. 57–65.

Nguyen, P., Leray, V., Diez, M., Serisier, S., Le Bloc'H, J., Siliart, B., *et al.* (2008) 'Liver lipid metabolism', *Journal of Animal Physiology and Animal Nutrition*, 92(3), pp. 272–283.

O'Hea, E. K. and Leveille, G. A. (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(3), pp. 338–344.

Pérez-Enciso, M., Clop, A., Noguera, J. L., Ovilo, C., Coll, A., Folch, J. M., *et al.* (2000) 'A QTL on pig chromosome 4 affects fatty acid metabolism : evidence from an Iberian by Landrace intercross', *Journal of animal science*, 78(10), pp. 2525–2531.

Price, N. L., Holtrup, B., Kwei, S. L., Wabitsch, M., Rodeheffer, M., Bianchini, L., *et al.* (2016) 'SREBP-1c/MicroRNA 33b Genomic Loci Control Adipocyte Differentiation.', *Molecular and cellular biology*, 36(7), pp. 1180–93.

Puig-Oliveras, A., Ballester, M., Corominas, J., Revilla, M., Estellé, J., Fernández, A. I., *et al.* (2014a) 'A co-association network analysis of the genetic determination of pig conformation, growth and fatness', *PLoS ONE*, 9(12), pp. 1–20.

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D., Hudson, N. J., *et al.* (2014b) 'Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition', *PLoS ONE*, 9(6).

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A. I., Folch, J. M. and Ballester, M.

(2016) 'Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat.', *Scientific reports*. Nature Publishing Group, 6(February), p. 31803.

R Core Team (2018) 'R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.'

Ramayo-Caldas, Y., Ballester, M., Fortes, M. R. S., Esteve-Codina, A., Castelló, A., Noguera, J. L., *et al.* (2014) 'From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine', *BMC Genomics*, 15(1).

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., *et al.* (2012a) 'Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition', *BMC Genomics*, 13(1).

Ramayo-Caldas, Y., Mercadé, A., Castelló, A., Yang, B., Rodríguez, C., Alves, E., *et al.* (2012b) 'Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross', *Journal of Animal Science*, 90(9), pp. 2883–2893.

Ramirez, C. M., Goedeke, L., Rotllan, N., Yoon, J.-H., Cirera-Salinas, D., Mattison, J. A., *et al.* (2013) 'MicroRNA 33 Regulates Glucose Metabolism', *Molecular and Cellular Biology*, 33(15), pp. 2891–2902.

Rayner, K. J., Suárez, Y., Dávalos, A., Parathath, S., Michael, L., Tamehiro, N., *et al.* (2010) 'MiR-33 Contributes to the Regulation of Cholesterol Homeostasis', *Science*, 328(5985), pp. 1570–1573.

Reddy, A. M., Zheng, Y., Jagadeeswaran, G., Macmil, S. L., Graham, W. B., Roe, B. A., *et al.* (2009) 'Cloning, characterization and expression analysis of porcine microRNAs', *BMC Genomics*, 10, pp. 1–15.

Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., Criado-Mesas, L., Castelló, A., Fernández, A. I., *et al.* (2018) 'Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions', *Scientific Reports*, 8(1), pp. 1–13.

Rocha, D. and Plastow, G. (2006) 'Commercial pigs: an untapped resource for human obesity research?', *Drug Discovery Today*, 11(11–12), pp. 475–477.

Shao, F., Wang, X., Yu, J., Jiang, H., Zhu, B. and Gu, Z. (2014) 'Expression of miR-33 from an SREBF2 intron targets the FTO gene in the chicken', *PLoS ONE*, 9(3), pp. 1–8.

Shen, W., Le, S., Li, Y. and Hu, F. (2016) 'SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation', *PLoS ONE*, 11(10), pp. 1–10.

Shimano, H. (2001) 'Sterol Regulatory Element-Binding Proteins (SREBPs) as Regulators of Lipid Metabolism', *Progress in Lipid Research*, (40), pp. 439–452.

Song, Z., Cooper, D. K. C., Cai, Z. and Mou, L. (2018) 'Expression and regulation profile of mature microRNA in the pig: Relevance to xenotransplantation', *BioMed Research International*. Hindawi, 2018.

Taniguchi, M., Nakajima, I., Chikuni, K., Kojima, M., Awata, T. and Mikawa, S. (2014) 'MicroRNA-33b downregulates the differentiation and development of porcine preadipocytes', *Molecular Biology Reports*, 41(2), pp. 1081–1090.

Timoneda, O., Balcells, I., Córdoba, S., Lló, A. C. and Sánchez, A. (2012) 'Determination of reference microRNAs for relative quantification in porcine tissues', *PLoS ONE*, 7(9).

Tsai, A. C., Romsos, D. R. and Leveille, G. A. (1975) 'Effect of Dietary Cholesterol on Hepatic Lipogenesis and Plasma Insulin and Free Fatty Acid Levels in Rats', *The Journal of Nutrition*, 105(7), pp. 939–945.

Webb, E. C. and O'Neill, H. A. (2008) 'The animal fat paradox and meat quality', *Meat Science*, 80(1), pp. 28–36.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., *et al.* (2008) 'Fat deposition, fatty acid composition and meat quality: A review', *Meat Science*, 78(4), pp. 343–358.

Wood, J. and Whittemore, C. (2007) *Pig Meat and Carcass Quality*, *Whittemore's Science and Practice of Pig Production*.

Xu, J., Nakamura, M. T., Cho, H. P. and Clarke, S. D. (1999) 'Sterol regulatory element binding protein-1 expression is suppressed by dietary polyunsaturated fatty acids. A

mechanism for the coordinate suppression of lipogenic genes by polyunsaturated fats', *Journal of Biological Chemistry*, 274(33), pp. 23577–23583.

Zhang, B. han, Shen, C. an, Zhu, B. wei, An, H. ying, Zheng, B., Xu, S. bo, *et al.* (2019) 'Insight into miRNAs related with glucometabolic disorder', *Biomedicine and Pharmacotherapy*, 111(51), pp. 657–665.

Supplementary Table 1. Primers used for *SREBF2*, *ACTB* and *TBP* gene expression quantification by qPCR.

Supplementary Table 2. List of genes analysed by qPCR in each tissue (Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017; Revilla *et al.*, 2018).

Supplementary Table 3. Pearson's correlations between miR-33a/b and mRNA target genes expression values. *P*-values are indicated in brackets and * means statistically significant (p-value < 0.05).

Supplementary Table 4. Pearson's correlation values between miR-33a and *SREBF2* gene expression, and between miR-33b and *SREBF1* gene expression.

PAPER IV

Unraveling porcine muscle gene expression regulators through expression genome-wide association studies

Lourdes Criado-Mesas[1], *et al*.

[1]Departament de Genòmica Animal, Centre de Recerca en Agrigenòmica (CRAG), CSIC-IRTA-UAB-UB, Barcelona, Spain.

E-mail: lourdes.criado@cragenomica.es

Manuscript in preparation

Abstract

Pig is one of the main sources of meat in the world, so the quality of its meat and nutritional values are gaining more interest. The main objective of this work was to study the *Longissimus dorsi* muscle transcriptomic profile of 132 Iberian x Duroc crossbreed pigs by RNA-Seq to identify potential muscle gene-expression regulators. The muscle gene expression data and the SNP genotypes obtained from the Axiom Porcine Genotyping Array (Affymetrix) were used to perform the expression genome-wide association studies and the expression quantitative trait *loci* (eQTL) mapping. A total of 291 genes showed significant associations with expression-SNPs and 324 eQTLs regions were identified, being 247 *cis*-eQTLs and 77 *trans*-eQTLs, and some hotspots were noticed. The two most significant associations were found for *HGFAC* and *HUS1* genes. The main representative processes were metabolic pathways, the most cellular component was the membrane and the top molecular function was ion binding. The functional analysis of the significant associated genes identified the top three canonical pathways: Granzyme B signaling, glutathione-mediated detoxification and NRF2-mediated Oxidative Stress Response. Moreover, *HNF4A*, *KLF3, E2F4,* the miR-483 and *RORC* were reported as the main transcription factors, nuclear receptors or miRNAs involved in the muscle gene expression regulation. Our results increase the knowledge of the genomic architecture of the pig skeletal muscle, but further analyses are needed to better identify potential candidate genes involved in the muscle gene expression regulation.

Introduction

In recent years, meat production is rising due the increase of the human population. The pig is an important livestock animal because it is one of the main sources of meat in the world. Meat is considered an important source of nutrients, although a high consumption can increase the risk of some types of chronic diseases (Godfray *et al.*, 2018). On the other hand, consumers are increasingly more concerned about healthy and high-quality meats. For instance, muscle growth and fat deposition are related to pork meat quality and nutritional values, and are considered the two most important traits of economic interest on pig production (Wood *et al.*, 2004, 2008). The muscle, together with liver and adipose tissue, are the main organs involved in the regulation of lipid metabolism. Muscle is a reservoir of amino acids, which are necessary for protein synthesis or energy production, and glucose (Meyer *et al.*, 2002).

Differential gene expression patterns in a specific tissue can explain the molecular bases of phenotypic differences among animals. The development of next generation sequencing technologies has provided new tools for both gene-expression profiling and transcriptome characterization. The RNA-Sequencing (RNA-Seq) method is based on the sequencing of RNA molecules present in a given sample. The obtained counts corresponding to each transcript can be used for quantification and the sequences can be mapped to the genome for their annotation. The transcriptome analyses allow not only the analysis of gene expression variation but also the identification of new isoforms, splicing events, and different promoter and polyadenylation signal usage.

In pigs, RNA-Seq was widely used to identify genes involved in fat deposition and meat quality. Previous studies on the porcine muscle transcriptome identified differentially expressed genes affecting backfat thickness (Xing *et al.*, 2016), intramuscular fat (Puig-Oliveras *et al.*, 2014; Cardoso, Cánovas and Amills, 2017; Chen *et al.*, 2017; Muñoz *et al.*, 2018; González-Prendes *et al.*, 2019), drip los (Heidt *et al.*, 2013; Li *et al.*, 2016) and lipid metabolism (Steibel *et al.*, 2011), among other traits. Moreover, the identification of genomic loci regulating the expression of genes through expression quantitative trait loci (eQTL) mapping may help to the identification of candidate genes, causal variants, and molecular pathways associated with phenotypic traits in pigs.

The main objective of this work was to study the muscle transcriptome of 132 Iberian x Duroc crossbreed pigs (BC1_DU) by RNA-Seq and to identify muscle gene-expression regulators.

Material and methods

Pig population

A total of 132 animals from an experimental backcross named BC1_DU (25% Iberian and 75% Duroc) were studied. Animal care and procedures were carried out following the Spanish Policy for Animal Protection RD1201/05 and the European Union Directive 86/609 about the protection of animals used in experimentation. All animals were maintained under the same intensive conditions and fed ad libitum with a cereal-based commercial diet. After, they were slaughtered in a commercial abattoir according to the institutional and national guidelines for the Good Experimental Practices and approved by the Ethical Committee of the Institution (IRTA – Institut de Recerca i Tecnologia Agroalimentàries), and sample tissues were collected in liquid nitrogen and stored at -80°C until analysis.

RNA isolation and sequencing

Total RNA extraction from muscle tissue was performed with the RiboPure kit (Ambion), following the manufacturer protocol. The RNA quantification and purity were assessed with a NanoDrop ND-1000 spectrophotometer (NanoDrop products) and RNA integrity was evaluated by Agilent Bioanalyzer-2100 (Agilent Technologies), and samples with a RIN greater than 7 were used for the study. Sequencing was performed at the CNAG institute (Centro Nacional de Análisis Genómico, Barcelona, España). Libraries for 132 samples were generated using the TruSeq Stranded mRNA kit (Illumina), following the manufacturer's recommendations, and sequenced on the Illumina HiSeq 3000/4000instrument, generating a mean of 44.2 million of 75 bp paired-end reads per sample.

Mapping and annotation of reads

We run MultiQC (Ewels *et al.*, 2016) for the quality control. RNA-seq reads were mapped by using the STAR software (Dobin *et al.*, 2013) with default parameters to the reference genome assembly *Sscrofa* 11.1 and to the annotation database Ensemble Genes 97. Transcripts were assembled and quantified by HTSeq (Anders, Pyl and Huber, 2015).

Gene expression quantification

Data pre-processing and quality control were performed with the EdgeR (Robinson, McCarthy and Smyth, 2009) and Limma (Ritchie *et al.*, 2015) R packages. First unexpressed genes were filtered and retained genes having more than 1 read per million in at least 25% of samples. Finally, a total of 11.413 genes were considered to be expressed in the muscle samples and further analyzed. Expressed gene counts were normalized using the log counts per million (logCPM) with the Limma-trend approach (Law *et al.*, 2014).

Genotyping

Genomic DNA was extracted from diaphragm tissue using the phenol-chloroform method (Sambrook, Fritsch and Maniatis, no date). Animals were genotyped using the Axiom Porcine Genotyping Array (Affymetrix) and only SNPs mapping against the *Sscrofa* 11.1 assembly were used. Moreover, SNPs were filtered using the PLINK software (Purcell *et al.*, 2007) to remove markers with a minor allele frequency (MAF) lower than 15%, SNPs with more than 1% of missing genotypes and a Hardy-Weinberg disequilibrium (HWD) less than 0.000001. Finally, a total of 308.512 SNPs distributed along all chromosomes were used for the analysis.

Expression Genome-wide association studies

Genomic association studies between each gene expression measure and SNPs genotypes (eGWAS) were performed through a linear mixed model using GEMMA software(Zhou and Stephens, 2012):

$y = W\alpha + x\beta + u + \varepsilon$; $u \sim MVNn(0, \lambda\tau^{-1}K)$, $\varepsilon \sim MVNn(0, \tau^{-1}In)$,

in which: y was the vector of phenotypes for n individuals; W is a matrix nxc of covariables (fixed effects) that includes a column of ones, sex (2 levels) and slaughtering batch (5 levels); α is a c vector with corresponding coefficients, including the intercept; x is a n vector with the marker genotypes; β is the size of the marker effect, u is an n vector of random effects (additive genetic effects), ε is an n vector of errors. The random effects vector is assumed to follow a normal multivariate n-dimensional distribution (MVNn) where τ-1 is the variance of residual errors; λ is the quotient between the two components of variance; K is an nxn matrix of kin calculated from the SNPs. The vector of errors is assumed to follow a distribution MVNn, where In is an nxn identity matrix.

GEMMA software calculates from the Wald statistical test the p-value for each SNP comparing the null hypothesis that the SNP has no effect versus the alternative hypothesis that the SNP effect is different from zero.

After, multiple testing corrections were performed at two levels using the FDR (False Discovery Rate) method of Benjamini and Hochberg (Benjamini and Hochberg, 1995) with the function *p.adjust* of R. First, for each gene individually, *p*-values of the associations between the gene expression variations and all SNPs were corrected by the FDR procedure at 5% (local FDR). Secondly, we assessed the significance of all associations of all genes together and a second significance threshold was calculated by FDR at 5% for all the *p*-values of the associations (global FDR). Only associations with a *p*-value lower than these established thresholds were considered significant. The number of significant SNPs belonging to the same interval was considered among associated SNPs less than 10 Mb apart.

Gene annotation

Genomic eQTLs intervals were annotated considering ±1 Mb around the candidate chromosomal regions. The extraction of the genes contained in the associated regions was performed with the BioMart (Smedley *et al.*, 2015) tool from the Ensembl project (www.ensembl.org; release 92) using the *Sscrofa* 11.1 reference assembly. Functional predictions of the significant SNPs was performed with the Ensembl Genes 97

Database using the Variant Effect Predictor (McLaren *et al.*, 2010) tool from the Ensembl project.

Gene functional classification

The WebGestalt (Zhang, Kirov and Snoddy, 2005) program was used to perform the functional enrichment analysis of genes found significantly associated in the eGWAS studies. Ingenuity Pathway Analysis software (IPA; Ingenuity Systems) and the Core Analysis function was used to identify the main biological functions and pathways of genes mapped within the eQTLs regions. In addition, ClueGO plug-in (Bindea *et al.*, 2009) was used to integrate and cluster the genes regarding their Gene Ontology and KEGG pathway.

Results and discussion

Identification of expression QTLs in the porcine skeletal muscle

We performed an eGWAS by combining the muscle gene expression data, measured by RNA-Seq, and the SNP genotypes obtained from the Axiom Porcine Genotyping Array (Affymetrix), in a total of 132 BC1_DU animals. The eGWAS identified 39.428 expression-SNPs (eSNPs) located in all *Sus scrofa* chromosomal regions and associated with the expression of 291 genes (FDR<0.05). In addition, chromosomes SSC2, SSC6, SSC7, SSC8, SSC9, SSC13, and SSC15 presented a higher number of significant associated eSNPs, while chromosomes SSC5, SSC11, and SSC18 showed a lower number of significant associated eSNPs. The identified significant associated eSNPs were classified depending on their location, as *cis*-eSNPs if they were located within 1 Mb of the analyzed gene, and as *trans*-eSNP if they were located elsewhere in the genome.

A total of 324 expression quantitative trait loci (eQTLs) regions were identified, being 247 *cis*-acting eQTLs and 77 *trans*-acting eQTLs (Figure 1). The analysis of eQTL genomic locations revealed that both *cis* and *trans* associations were widely distributed on all chromosomes.

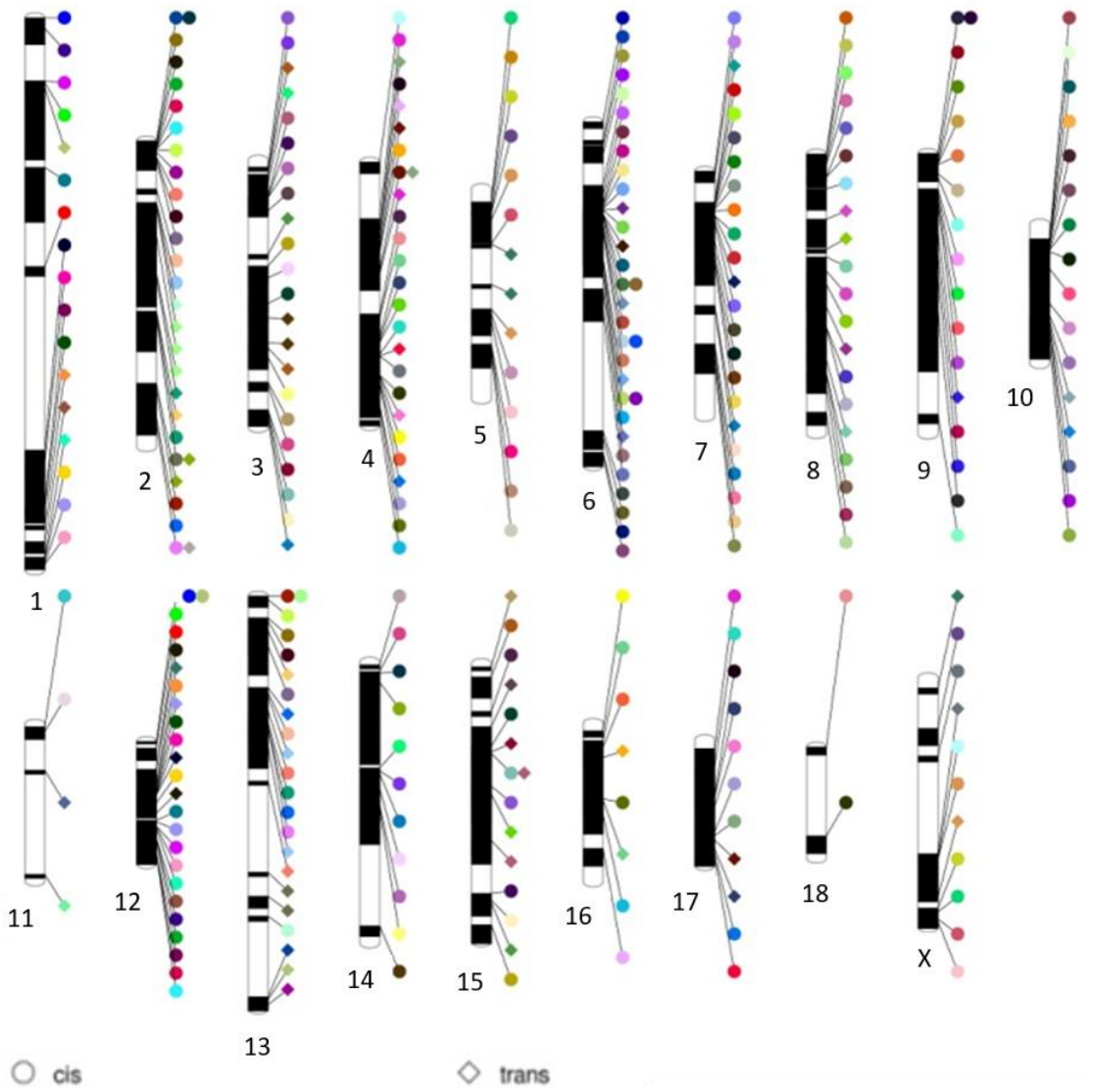


Figure 1. PhenoGram plot representing the distribution of the eQTLs identified along all pig chromosomes. The colour indicates each different gene and the shape indicates *cis* or *trans*-acting eQTL.

Cis-eQTL regions are described to alter the expression of nearby genes, while *trans*-eQTL regions are associated with the expression of remote genes, usually located on different chromosomes (Cheung and Spielman, 2009). *Trans*-acting effects are often weaker than *cis* effects, so true *trans*-eQTL are more difficult to detect than *cis*-eQTLs. Moreover, sometimes one single location is associated with the regulation of multiple genes and it is defined as a *trans*-eQTL hotspot (Schadt *et al.*, 2003; Pierce *et al.*, 2014).

In the current study, 7 eQTL hotspot regions were identified for 17 different genes (Table 1).

Interval	Gene	Chr.	Start-End Position (Mb)	SNPs	Candidate Genes
1	*ENSSSCG00000016886*; *ENSSSCG00000027013*; *ENSSSCG00000027013*; *ENSSSCG00000014101* (*AP3B1*)	2	31.7-82.1	673	*KLF2*, *SMARCA4*, and *TEAD1*
2	*ENSSSCG00000033693* (*SLC10A5*), *ENSSSCG00000014875* (*CAPN5*)	4	39.2-44.7	18	*GDF6*
3	*ENSSSCG00000033100*; *ENSSSCG00000033790*	4	106-122	165	
4	*ENSSSCG00000031204*; *ENSSSCG00000003022* (*TMEM145*)	6	45.6-70	1251	*ZFP36*
5	*ENSSSCG00000026140*; *ENSSSCG00000035593*; *ENSSSCG00000035756*	13	202.2-209.2	26	
6	*ENSSSCG00000015854* (*OCA2*); *ENSSSCG00000034866*	15	44.6-80.7	891	
7	*ENSSSCG00000002529*; *ENSSSCG00000007337* (*CTNNBL1*)	17	50.6-63.9	191	

Table 1. Significant *trans*-eQTLs for the hotspot regions found. Start and end positions refer to the eQTL interval and are based on *Sscrofa* 11.1 assembly. Gene annotation

was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval.

The first hotspot *trans*-eQTL region was located on SSC2 (31.7-82.1 Mb) and three transcription factors related to muscle transcriptome regulation were mapped: *KLF2*, which is highly expressed in endothelial cells and lymphocytes (Dekker *et al.*, 2002), has no clear role in skeletal muscle, *SMARCA4* which is involved in early transcription of genes and muscle differentiation during myogenesis (Albini *et al.*, 2015), and *TEAD1*, a key transcription factor for muscle development activating multiple genes involved in cell proliferation and differentiation pathways (Hsu *et al.*, 1996). A *trans*-eQTL hotspot located on SSC4 (34.95-44.6 Mb) was associated with the expression of two genes: *SLC10A5*, which is a member of the bile acid transporters, and *CAPN5*, which is involved in signal transduction in different cellular processes. In this hotspot region, the *GDF6* gene was mapped and has been reported to encode for a ligand of the TFG-beta receptors, leading to the recruitment and activation of SMAD family transcription factors. The *trans*-eQTL hotspot located on SSC6 was associated with the expression of two genes and the *ZFP36* gene was mapped, and a member of his family, *ZFP36L2* gene, was described as a regulator of genes related to cell death or apoptosis pathways (Ponsuksili *et al.*, 2015).

In other studies, hotspot regions were potentially identified as key regulators of many downstream genes and pathways (Breitling *et al.*, 2008). For instance, different hotspot regions were identified regulating genes involved in different production traits, such as growth (Steibel *et al.*, 2011; Ponsuksili *et al.*, 2012), fatness and fatty acid composition (Ponsuksili *et al.*, 2011; Steibel *et al.*, 2011; Cánovas *et al.*, 2012; Muñoz *et al.*, 2013b; Martínez-Montes *et al.*, 2017; González-Prendes *et al.*, 2019) and meat quality (Ponsuksili *et al.*, 2008, 2010, 2014; Wimmers, Murani and Ponsuksili, 2010; Steibel *et al.*, 2011; Heidt *et al.*, 2013; Muñoz *et al.*, 2013a; Pena *et al.*, 2013; Manunza *et al.*, 2014; González-Prendes *et al.*, 2017), by using microarrays expression data. On the contrary, only one work reported hotspot regions in the pig muscle using RNA-Seq data (Velez-Irizarry *et al.*, 2019).

Then, we applied a stringent double multiple testing correction, at both local and global level, the hundreds of eQTLs identified in our study are expected to be just a

small part of the total number. Hence, eQTLs with small effects were likely missed due the lack of statistical power to detect them.

The most significant associations found in the eGWAS were for the *HGFAC* gene expression and the *AX-116337078* polymorphism (*p*-value = $4.4.x10^{-58}$), and the *HUS1* gene and the *AX-116695068* polymorphism (*p*-value = $1.17x10^{-55}$) (Figure 2).

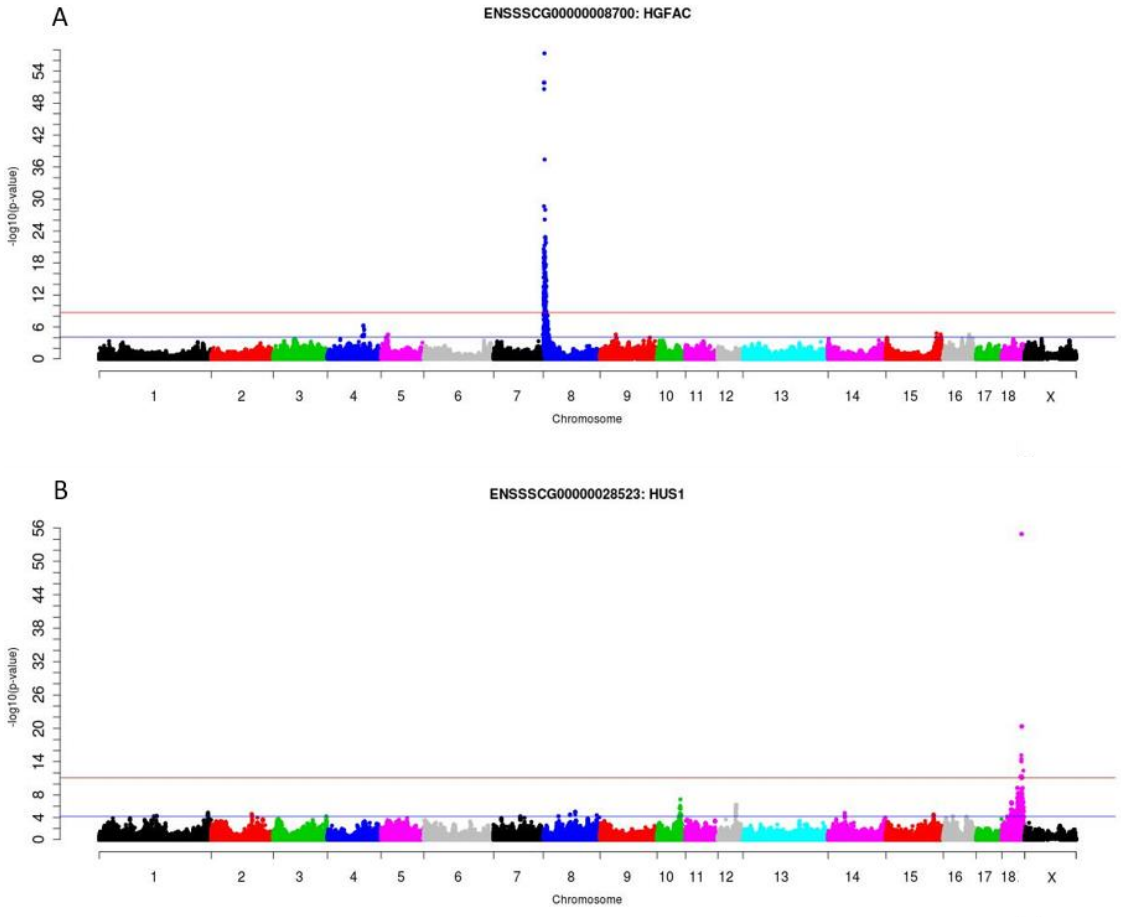


Figure 2. GWAS plot of muscle *HGFAC* (A) and *HUS1* (B) gene expression. Chromosome positions in Mb based on *Sscrofa* 11.1 assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. Horizontal lines represent the genome-wide significance level (local FDR-based q-value < 0.05 corresponds to blue line, and global FDR-based q-value < 0.05 to red line).

The *HGF Activator* (*HGFAC*) gene encodes for a proteinase and has been reported to be involved in tissue regeneration and repair (Fukushima *et al.*, 2018) and the *HUS1*

Checkpoint Clamp Component is involved in the cell cycle arrest in response to DNA damage (O'Connell, Walworth and Carr, 2000).

Finally, a total of 20,764 genes were mapped in the 324 different eQTL regions. The widespread location of the eQTLs and the high number of genes mapped in their regions indicates that more restrictive parameters should be applied in the definition of the intervals and in the mapping of the genes in the eQTL regions.

Functional analysis

A functional analysis was done with the 291 genes significantly associated with the gene expression in the eGWAS studies. The top biological processes, cellular components and molecular functions were analysed using the WebGestalt (Zhang, Kirov and Snoddy, 2005) program and results are presented in Figure 3. Most of the genes were functionally unclassified in the three categories. Metabolic pathways were the most representative process; membrane was the most present cellular component and ion binding the top molecular function category.

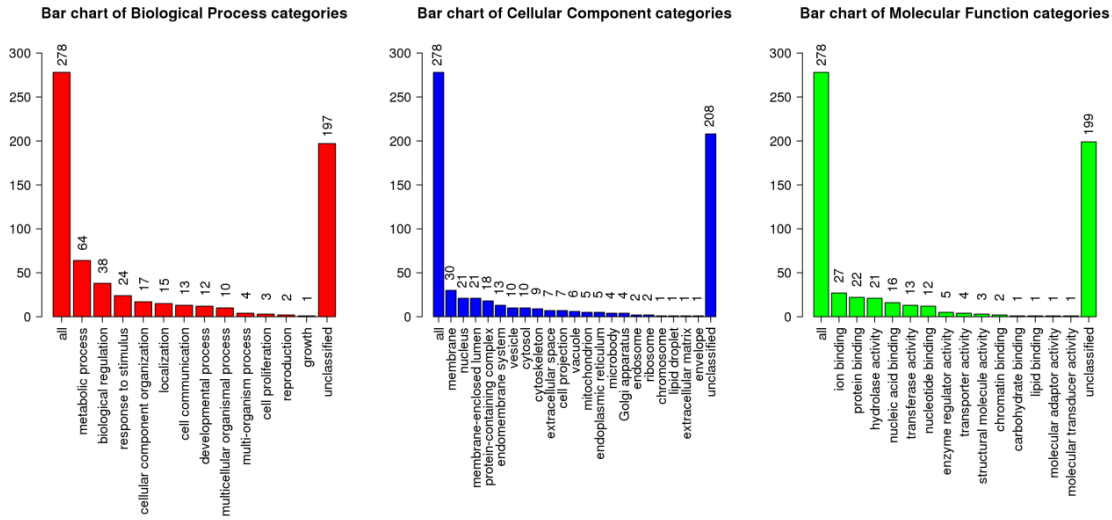


Figure 3. Bar chart representing the biological process, the cellular component and the molecular function categories of the 291 significantly associated genes identified in the eGWAS studies.

To have a comprehensive functional analysis of the genes significantly associated with the gene expression in the eGWAS studies, we used the Ingenuity Pathway Analysis (Krämer *et al.*, 2014) program. The top canonical pathways overrepresented according to IPA were related with Granzyme B signaling (3 genes, *p*-value=$7.01x10^{-04}$), glutathione-mediated detoxification (3 genes, *p*-value=$5.44x10^{-03}$) and NRF2-mediated Oxidative Stress Response (7 genes, *p*-value=$5.60x10^{-03}$). On the other hand, the most relevant molecular and cellular functions were cell morphology (26 molecules, *p*-value range= $2.23x10^{-02}$ – $1.25x10^{-04}$), cellular assembly and organization (23 molecules, *p*-value range= $2.23x10^{-02}$ – $1.25x10^{-04}$), cellular function and maintenance (22 molecules, *p*-value range= $2.23x10^{-02}$ – $1.25x10^{-04}$), nucleic acid metabolism (7 molecules, *p*-value range= $2.23x10^{-02}$ – $3.73x10^{-04}$), and small molecule biochemistry (42 molecules, *p*-value range= $2.23x10^{-02}$ – $3.73x10^{-04}$).

In addition, we analyzed the possible implication of different transcription factors, miRNAs or nuclear receptors in the gene expression variation. The most important transcription factor found was the *HNF4A* gene (Figure 4), which has been related with the regulation of several genes and has been involved in the insulin pathway and the IMF accumulation (Ayuso *et al.*, 2016). In particular, it has been reported that the *HNF4A/PPARGC1A* pathways are activating gluconeogenic genes in the liver during the fasting time and *SREBP1* is inhibiting the genes during the fed state (Yamamoto *et al.*, 2004). Moreover, a polymorphism located near the *HNF4A* gene influences the activity of the enzyme isocitrate dehydrogenase which is related to the oxidative state of muscle fibers to catabolize fatty acids (Sevane *et al.*, 2013).

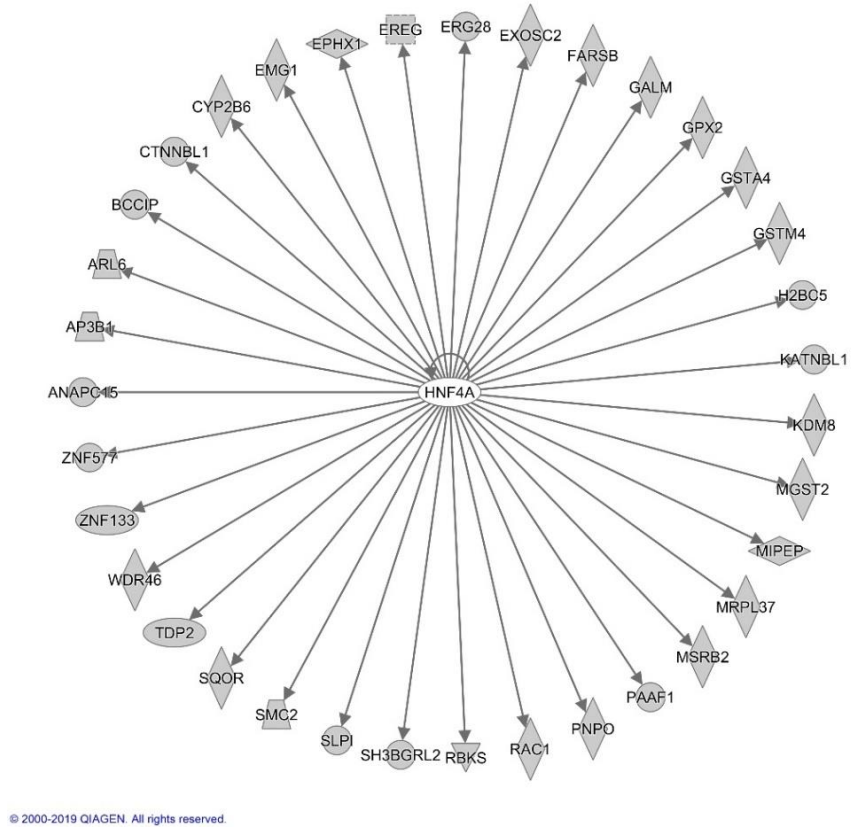


Figure 4. Network generated by IPA of *HNF4A* gene and their target genes.

Other relevant transcription factors identified were *Kruppel Like Factor 3* (*KLF3*) and *E2F Transcription Factor 4* (*E2F4*), as well as the *mir-483* and the nuclear receptor *RAR Related Orphan Receptor C* (*RORC*). The *KLF3* transcription factor has been identified to regulate muscle-specific gene expression and interacts with the serum response factor, which is a crucial transcription factor for the cardiac, skeletal and smooth muscle gene expression (Himeda *et al.*, 2010). On the other hand, the *E2F4* is a transcription factor that has been reported to interact with the *ACSL1* gene in bovines, which plays an important role in fatty acid transport and degradation in addition to lipid synthesis (Zhao *et al.*, 2016). The miR-483, which is located in the second intron of the *IGF2* gene, has been reported to be co-expressed with its host gene and related with the IGF2 growth factor, which suggest its possible implication in the regulation of the metabolism (Ma *et al.*, 2011). Finally, *RORC* forms heterodimers with *PPARG* and *RXRG* transcription factors, it is highly expressed in skeletal muscle and has been identified as a candidate gene for type 2 diabetes (Wang *et al.*, 2003). Between the *RORC* target

genes are *CYP4F8* and *CYP2B6*, which are involved in the synthesis of lipids, *GSTM4*, a member of the glutathione metabolism, *RDH16*, a gene implicated in the retinoic acid pathway and the *SLC25A24* gene, which is a mitochondrial carrier.

In summary, the eGWAS revealed both *cis* and *trans*-eQTLs for the gene expression in the pig muscle and the significant associated genes were mainly involved in metabolic pathways. We proposed the *HNF4A*, *KLF3*, *E3F4* and *RORC* transcription factors and nuclear receptors as candidate genes involved in the regulation of the gene expression in the pig muscle. Altogether, the identified genes and their functions and pathways increase our knowledge of the genomic architecture of the pig muscle. However, a refined eQTL interval definition and gene mapping should be done, as well as network studies to identify potential candidate genes involved in the muscle gene expression regulation.

References

Albini, S., Toto, P. C., Agnese, A. D., Malecova, B., Cenciarelli, C., Felsani, A., *et al.* (2015) 'Brahma is required for cell cycle arrest and late muscle gene expression during skeletal myogenesis', 16(8), pp. 1037–1050.

Anders, S., Pyl, P. T. and Huber, W. (2015) 'HTSeq-A Python framework to work with high-throughput sequencing data', *Bioinformatics*, 31(2), pp. 166–169.

Ayuso, M., Fernández, A., Núñez, Y., Benítez, R., Isabel, B., Fernández, A. I., *et al.* (2016) 'Developmental stage, muscle and genetic type modify muscle transcriptome in pigs: Effects on gene expression and regulatory factors involved in growth and metabolism', *PLoS ONE*, 11(12), pp. 1–33.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society*, 57(1), pp. 289–300.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., *et al.* (2009) 'ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology

and pathway annotation networks', *Bioinformatics*, 25(8), pp. 1091–1093.

Breitling, R., Li, Y., Tesson, B. M., Fu, J., Wu, C., Wiltshire, T., *et al.* (2008) 'Genetical Genomics : Spotlight on QTL Hotspots', *PLoS Genetics*, 4(10), pp. 2–5.

Cánovas, A., Pena, R. N., Gallardo, D., Ramírez, O., Amills, M. and Quintanilla, R. (2012) 'Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population', *PLoS ONE*, 7(4), pp. 1–12.

Cardoso, T. F., Cánovas, A. and Amills, M. (2017) 'RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles', *Scientific Reports*, pp. 1–9.

Chen, W., Fang, G. feng, Wang, S. dong, Wang, H. and Zeng, Y. qing (2017) 'Longissimus lumborum muscle transcriptome analysis of Laiwu and Yorkshire pigs differing in intramuscular fat content', *Genes and Genomics*, 39(7), pp. 759–766.

Cheung, V. G. and Spielman, R. S. (2009) 'Genetics of human gene expression: Mapping DNA variants that influence gene expression', *Nature Reviews Genetics*, 10(9), pp. 595–604.

Dekker, R. J., Soest, S. Van, Fontijn, R. D., Salamanca, S., Groot, P. G. De, Vanbavel, E., *et al.* (2002) 'Prolonged fluid shear stress induces a distinct set of endothelial cell genes , most specifically lung Kruppel-like factor (KLF2)', *Blood*, 100(5), pp. 1689–1698.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) 'MultiQC: Summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19), pp. 3047–3048.

Fukushima, T., Uchiyama, S., Tanaka, H. and Kataoka, H. (2018) 'Hepatocyte growth factor activator: A proteinase linking tissue injury with repair', *International Journal of Molecular Sciences*, 19(11), pp. 1–11.

Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., *et al.* (2018)

'Meat consumption, health, and the environment', *Science*, 361(6399).

González-Prendes, R., Quintanilla, R., Cánovas, A., Manunza, A., Cardoso, T. F., Jordana, J., *et al.* (2017) 'Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits', *Scientific Reports*, 7, pp. 1–9.

González-Prendes, R., Quintanilla, R., Mármol-Sánchez, E., Pena, R. N., Ballester, M., Cardoso, T. F., *et al.* (2019) 'Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine gluteus medius and longissimus dorsi muscles', *BMC Genomics*. BMC Genomics, 20(1), pp. 1–18.

Heidt, H., Cinar, M. U., Uddin, M. J., Looft, C., Jüngst, H., Tesfaye, D., *et al.* (2013) 'A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population', *Mammalian Genome*, 24(9–10), pp. 416–426.

Himeda, C. L., Ranish, J. A., Pearson, R. C. M., Crossley, M. and Hauschka, S. D. (2010) 'KLF3 Regulates Muscle-Specific Gene Expression and Synergizes with Serum Response Factor on KLF Binding Sites', *Molecular and Cellular Biology*, 30(14), pp. 3430–3443.

Hsu, D. K. W., Guo, Y., Alberts, G. F., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., *et al.* (1996) 'Identification of a Murine TEF-1-related Gene Expressed after Mitogenic Stimulation of Quiescent Fibroblasts and during Myogenic Differentiation', *Journal of biological chemistry*, 271(23), pp. 13786–13795.

Krämer, A., Green, J., Pollard, J. and Tugendreich, S. (2014) 'Causal analysis approaches in ingenuity pathway analysis', *Bioinformatics*, 30(4), pp. 523–530.

Law, C. W., Chen, Y., Shi, W. and Smyth, G. K. (2014) 'Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts', *Genome Biology*, 15(2), pp. 1–17.

Li, B., Dong, C., Li, P., Ren, Z., Wang, H., Yu, F., *et al.* (2016) 'Identification of candidate genes associated with porcine meat color traits by genome-wide transcriptome analysis', *Scientific Reports*, 6.

Ma, N., Wang, X., Qiao, Y., Li, F., Hui, Y., Zou, C., *et al.* (2011) 'Coexpression of an intronic microRNA and its host gene reveals a potential role for miR-483-5p as an IGF2

partner', *Molecular and Cellular Endocrinology*, 333(1), pp. 96–101.

Manunza, A., Casellas, J., Quintanilla, R., González-Prendes, R., Pena, R. N., Tibau, J., *et al.* (2014) 'A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants', *BMC Genomics*, 15(1), pp. 1–12.

Martínez-Montes, A. M., Muiños-Bühl, A., Fernández, A., Folch, J. M., Ibáñez-Escriche, N. and Fernández, A. I. (2017) 'Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics', *Mammalian Genome*, 28(3–4), pp. 130–142.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) 'Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor', *Bioinformatics*, 26(16), pp. 2069–2070.

Meyer, C., Dostou, J. M., Welle, S. L. and Gerich, J. E. (2002) 'Role of human liver, kidney, and skeletal muscle in postprandial glucose homeostasis', *American Journal of Physiology-Endocrinology and Metabolism*, 282(2), pp. E419–E427.

Muñoz, M., García-Casco, J. M., Caraballo, C., Fernández-Barroso, M. Á., Sánchez-Esquiliche, F., Gómez, F., *et al.* (2018) 'Identification of Candidate Genes and Regulatory Factors Underlying Intramuscular Fat Content Through Longissimus Dorsi Transcriptome Analyses in Heavy Iberian Pigs', *Frontiers in Genetics*, 9, pp. 1–16.

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

Muñoz, M., Rodríguez, M. C., Alves, E., Folch, J. M., Ibañez-Escriche, N., Silió, L., *et al.* (2013b) 'Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data', *BMC Genomics*, 14(1).

O'Connell, M. J., Walworth, N. C. and Carr, A. M. (2000) 'The G2-phase DNA-damage checkpoint', *Trends in Cell Biology*, 10(7), pp. 296–303.

Pena, R. N., Noguera, J. L., Casellas, J., Díaz, I., Fernández, A. I., Folch, J. M., *et al.* (2013) 'Transcriptional analysis of intramuscular fatty acid composition in the longissimus thoracis muscle of Iberian × Landrace back-crossed pigs', *Animal Genetics*, 44(6), pp. 648–660.

Pierce, B. L., Tong, L., Chen, L. S., Rahaman, R., Argos, M., Jasmine, F., *et al.* (2014) 'Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians', *PLoS Genetics*, 10(12).

Ponsuksili, S., Du, Y., Murani, E., Schwerin, M. and Wimmers, K. (2012) 'Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle', *Genetics*, 192(3), pp. 1109–1122.

Ponsuksili, S., Jonas, E., Murani, E., Phatsara, C., Srikanchai, T., Walz, C., *et al.* (2008) 'Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle', *BMC Genomics*, 9, pp. 1–14.

Ponsuksili, S., Murani, E., Brand, B., Schwerin, M. and Wimmers, K. (2011) 'Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model', *Journal of Lipid Research*, 52(4), pp. 668–678.

Ponsuksili, S., Murani, E., Schwerin, M., Schellander, K. and Wimmers, K. (2010) 'Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits', *BMC Genomics*, 11(1).

Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M. and Wimmers, K. (2014) 'Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis', *International Journal of Biological Sciences*, 10(3), pp. 327–337.

Ponsuksili, S., Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., *et al.* (2015) 'Identification of common regulators of genes in co-expression networks affecting muscle and meat properties', *PLoS ONE*, 10(4), pp. 1–18.

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D.,

Hudson, N. J., *et al.* (2014) 'Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition', *PLoS ONE*, 9(6).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., *et al.* (2007) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., *et al.* (2015) 'Limma powers differential expression analyses for RNA-sequencing and microarray studies', *Nucleic Acids Research*, 43(7), p. e47.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2009) 'edgeR: A Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26(1), pp. 139–140.

Sambrook, J., Fritsch, E. and Maniatis, T. (no date) *Molecular Cloning: A Laboratory Manual 2nd edn. Cold Spring Harbor, N.Y.*

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., *et al.* (2003) 'Genetics of gene expression surveyed in maize, mouse and man', *Nature*, 422(6929), pp. 297–302.

Sevane, N., Armstrong, E., Cortés, O., Wiener, P., Wong, R. P. and Dunner, S. (2013) 'Association of bovine meat quality traits with genes included in the PPARG and PPARGC1A networks', *Meat Science*, 94(3), pp. 328–335.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., *et al.* (2015) 'The BioMart community portal: An innovative alternative to large, centralized data repositories', *Nucleic Acids Research*, 43(W1), pp. W589–W598.

Steibel, J. P., Bates, R. O., Rosa, G. J. M., Tempelman, R. J., Rilington, V. D., Ragavendran, A., *et al.* (2011) 'Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs', *PLoS ONE*, 6(2).

Velez-Irizarry, D., Casiro, S., Daza, K. R., Bates, R. O., Raney, N. E., Steibel, J. P., *et al.* (2019) 'Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs 06 Biological Sciences 0604

Genetics', *BMC Genomics*. BMC Genomics, 20(1), pp. 1–19.

Wang, H., Chu, W., Das, S. K., Zheng, Z., Hasstedt, S. J. and Elbein, S. C. (2003) 'Molecular screening and association studies of retinoid-related orphan receptor γ (RORC): A positional and functional candidate for type 2 diabetes', *Molecular Genetics and Metabolism*, 79(3), pp. 176–182.

Wimmers, K., Murani, E. and Ponsuksili, S. (2010) 'Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat performance in pigs', *Briefings in Functional Genomics and Proteomics*, 9(3), pp. 251–258.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., *et al.* (2008) 'Fat deposition, fatty acid composition and meat quality: A review', *Meat Science*, 78(4), pp. 343–358.

Wood, J. D., Richardson, R. I., Nute, G. R., Fisher, A. V., Campo, M. M., Kasapidou, E., *et al.* (2004) 'Effects of fatty acids on meat quality: A review', *Meat Science*, 66(1), pp. 21–32.

Xing, K., Zhu, F., Zhai, L., Chen, S., Tan, Z., Sun, Y., *et al.* (2016) 'Identification of genes for controlling swine adipose deposition by integrating transcriptome, whole-genome resequencing, and quantitative trait loci data', *Scientific Reports*, 6, pp. 1–10.

Yamamoto, T., Shimano, H., Nakagawa, Y., Ide, T., Yahagi, N., Matsuzaka, T., *et al.* (2004) 'SREBP-1 Interacts with Hepatocyte Nuclear Factor-4α and Interferes with PGC-1 Recruitment to Suppress Hepatic Gluconeogenic Genes', *Journal of Biological Chemistry*, 279(13), pp. 12027–12035.

Zhang, B., Kirov, S. and Snoddy, J. (2005) 'WebGestalt: An integrated system for exploring gene sets in various biological contexts', *Nucleic Acids Research*, 33(2), pp. 741–748.

Zhao, Z. D., Zan, L. Sen, Li, A. N., Cheng, G., Li, S. J., Zhang, Y. R., *et al.* (2016) 'Characterization of the promoter region of the bovine long-chain acyl-CoA synthetase 1 gene: Roles of E2F1, Sp1, KLF15, and E2F4', *Scientific Reports*, 6(January), pp. 1–9.

Zhou, X. and Stephens, M. (2012) 'Genome-wide efficient mixed-model analysis for association studies', *Nature genetics*, 44(7), pp. 821–4.

General Discussion

Chapter 4

Traditional and pure pig breeds are valuable resources for meat production, but also for cultural, historical and environmental aspects. Modern systems of intensive livestock production are dominated by highly productive global breeds, in which the disappearance of genetic variation concerns breeding companies (Hulsegge *et al.*, 2019). For the improvement of growth, fatness and meat quality traits among other traits of high economic impact, it is crucial to design better strategies for genetic selection. Nowadays, pork meat consumption in the world is rising basically due the human population growth, but it is static or declining in high-income countries. In pork meat production two market trends are becoming increasingly important: meat quality and the nutritional characteristics of the product (Godfray *et al.*, 2018).

Over the years, the genetic basis of meat quality traits has been studied and resulted in the characterization of several QTLs providing new insights about the genetic architecture of pork complex traits. Development of new genomic tools such microarrays, high-throughput SNP chips and NGS technologies provided massive results and allowed us to approach the identification of causal genes and mutations. In the IBMAP experimental population, several *loci* associated with growth, fatness and meat quality traits have been identified by QTL mapping and GWAS approaches, microarrays, RNA-Seq and systems genetics among others. Several positional candidate genes located in major QTLs were further analysed to identify causal polymorphisms, and genetic and functional validation experiments of these genetic variants were also performed.

This PhD thesis focused on the study of the genetic basis of fatty acid composition, an important determinant of meat quality, to identify and analyse candidate genes and mutations using molecular genetics and genomic technologies. For this purpose, the expression of 45 lipid-related genes in the muscle of three experimental backcrosses was analysed. After, we studied different candidate genes associated with fatty acid composition: *IGF2*, *mir-33a*, and *mir-33b*, in addition to the study of the *ELOVL6* gene. Finally, an RNA-Seq study of the muscle transcriptome of 132 animals from the BC1_DU population and eGWAS analysis for the identification of regulators of gene expression were performed. In the next sections, the main results obtained are discussed.

4.1. Candidate genes involved in fatty acid metabolism

Nowadays, genomic selection programs tend to incorporate biological information of the traits, which is critical for increasing accuracy in genomic predictions (Pérez-Enciso, Rincón and Legarra, 2015). In this line, we focused on candidate genes related to lipid metabolism in order to identify functional genetic variants that help to improve the porcine selection programmes.

First, a subset of 45 candidate lipid-related genes were analysed to better understand their gene expression regulation and the possible effect on fatty acid composition. Then, we focused on the *IGF2* gene, which is located in a QTL for muscle growth and fat deposition (SSC2). The *mir-33a* and *mir-33b*, which are located in intronic regions of the *SREBF* gene family, were also studied due their implication in lipid metabolism. Finally, an additional study was conducted on the *ELOVL6* gene, which is located in a QTL affecting fatty acid composition (SSC8).

4.1.1. Muscle gene expression study of 45 candidate genes for lipid metabolism

A strategy recently used to study the genetic architecture of complex traits is the detection of QTLs associated with gene expression levels (eQTLs), which results depend on the recombination frequency and the number of samples used (MacKay, Stone and Ayroles, 2009). Hence, gene expression values are considered quantitative traits and the eGWAS point to genetic variants associated with gene transcription levels. In the eGWAS studies, significant associations between the gene expression and genetic markers can be detected because the SNPs are in linkage disequilibrium with the causal mutation. In addition, genomic positions can be considered *cis*-acting when they are located close to the studied gene or *trans*-acting if they are located elsewhere. The identification of candidate genes within eQTL regions relies on the correct mapping and annotation of genes. The eQTL identification can deep also in the gene expression regulation mechanisms through gene network interactions. In general, genes involved in lipid metabolism are regulated at transcriptional level, and the study of the molecular mechanisms controlling its expression will help to understand the genetic basis of fatty acid composition in muscle tissue (Hausman *et al.*, 2009).

In the experimental IBMAP population, strong candidate genes affecting meat quality traits have been identified using QTL and GWAS, RNA-Seq and co-association network approaches. In previous studies of the liver, muscle and adipose tissue transcriptomes in the BC1_LD population, differentially-expressed genes were identified between two groups of extreme animals for fatty acid composition using RNA-Seq (Ramayo-Caldas *et al.*, 2012; Corominas *et al.*, 2013a; Puig-Oliveras *et al.*, 2014). In general, within the overrepresented pathways the PPAR signalling pathway was identified in the functional analysis of the three tissues analysed, and the differential expression of target genes for PPARs was observed. The obtained results supported that the variation in gene expression and its genetic basis could play an important role in the genetic determinism of these traits. In addition, Muñoz *et al*. (2013) performed an eQTL analysis using microarray gene expression data generated in *Longissimus dorsi* muscle samples for those genes that mapped within QTLs for fatty acid composition in the same population. Twelve eQTLs in SSC8, SSC11 and SSC17 for *BGLAP*, *ELOVL6*, *MGST2*, *PTPN11*, and *SEC13* were identified, but only an eQTL located on SSC8 for *MGST2* gene expression passed the FDR correction cut-off (0.2) (Muñoz *et al.*, 2013). The microarray technology provides an overview of gene expression in a whole genome scale but sometimes results are noisy or ambiguous (Spurgeon, Jones and Ramakrishnan, 2008). In another study, gene expression of 45 candidate genes related with lipid metabolism was analysed in the *Longissimus dorsi* muscle of 114 BC1_LD animals using the Fluidigm platform, a high-throughput microfluidic system that analyses gene expression by RT-qPCR (Puig-Oliveras *et al.*, 2016). The Fluidigm platform was selected because offers rapid, cost-effective and customizable arrays for a flexible moderate number of genes in several animals. It is also based on RT-qPCR, which confers a high sensibility and reproducibility, as well as is able to detect a large dynamic range of expression values (Spurgeon, Jones and Ramakrishnan, 2008).

Moreover, similar studies using Fluidigm platform to evaluate the mRNA expression levels of lipid-related candidate genes in liver and adipose tissue of 111 and 115 BC1_LD animals, respectively, were also performed in our group (Ballester *et al.*, 2017b; Revilla *et al.*, 2018). In the present work, *trans*-eQTLs showed to be more abundant than *cis*-eQTLs (Table 4.1). *Cis*-eQTLs are considered to contribute to the

variation of gene expression, while *trans*-eQTLs are considered regulatory hotspots because they generally regulate a large number of genes (Schadt *et al.*, 2003) and are tissue-specific (Gerrits *et al.*, 2009).

Table 4.1. Summary of the articles reported by our group describing the number of chromosomal regions associated with gene expression phenotypes in different tissues and populations. One and two of the eQTL regions showed both *cis* and *trans* effects, labelled as * and ** respectively.

Reference	Puig-Oliveras *et al.* 2016	Ballester *et al.* 2017	Revilla *et al.* 2018	Criado-Mesas *et al.* in revision
Tissue	Muscle	Liver	Adipose tissue	Muscle
Population	BC1_LD	BC1_LD	BC1_LD	BC1_LD, BC1_DU and BC1_PI
Associated genes	*ACSM5, CROT, FABP3, FOS, HIF1AN, IGF2, MGLL, NCOA1, PIK3R1, PLA2G12A* and *PPARA*	*CROT, CYP2U1, DGAT2, EGF, FABP1, FABP5, PLA2G12A* and *PPARA*	*ACSM5, ELOVL6, FABP4, FADS2* and *SLC27A4*	*ACSM5, ACSS2, ATF3, DGAT2, FOS* and *IGF2*
Total eQTLs	18*	7	19*	10**
***cis*-eQTLs**	3	2	3	2
***trans*-eQTLs**	16	5	17	10

In addition, gene-specific eQTLs have been performed for some candidate genes in the IBMAP cross, such as *ACSL4*, *APOA2*, *ELOVL6*, *FABP4*, *FABP5* and *IGF2* (Corominas *et al.*, 2012, 2013b; Ballester *et al.*, 2016, 2017a; Criado-Mesas *et al.*, 2019), using the RT-qPCR technology, to better understand their gene expression regulatory mechanisms.

In the present work, the expression and regulation of a selected set of 45 candidate genes involved in lipid metabolism in the porcine *Longissimus dorsi* muscle of 355

animals was studied, using the data generated in Puig-Oliveras *et al*. (2016) from BC1_LD and new data generated in the current study from the BC1_DU and BC1_PI populations. These three experimental backcrosses are based on the Iberian breed and the use of different populations tend to increase the accuracy of the genomic regions found due the reduction of the linkage disequilibrium in long regions (Goddard and Hayes, 2009). Moreover, increasing sample size is an advantage in the eGWAS analysis because improves the identification of associations if the genetic architecture of a trait has common variants of small effect. However, genetic heterogeneity can reduce the correlation between the phenotype and the genetic variants (Korte and Farlow, 2013). In addition, it is necessary to highlight that in the current study the newest *Sus Scrofa 11.1* assembly was used, in contrast to the *Sus Scrofa 10.2* assembly used in the work of Puig-Oliveras *et al*. (2016).

In the eGWAS analysis including the three populations (3BCs), we identified 186 expression-SNPs located in ten SSC regions and associated with the expression of *ACSM5*, *ACSS2*, *ATF3*, *DGAT2*, *FOS* and *IGF2* genes. All ten eQTL regions showed *trans* regulatory effects on gene expression and two of them were also identified in *cis* (*IGF2* and *ACSM5*). The *IGF2 cis*-eQTL region identified in the current study will be discussed below. On the other hand, the *ACSM5 cis*-eQTL region identified in the current work was also reported by Puig-Oliveras *et al.* (2016) in muscle, where the *ACSM5* proximal promoter region was amplified and sequenced in a subset of ten BC1_LD animals and three polymorphisms were identified. The proximal one (here known as *ACSM5.P*) was the most significantly associated with the *ACSM5* gene expression in the BC1_LD population, so we genotyped this candidate SNP in the BC1_DU and BC1_PI animals. Then, the *ACSM5.P* was the most significantly associated SNP with the muscle *ACSM5* gene expression in the 3BCs eGWAS and explained approximately the 40% of the phenotypic variance. Interestingly, Revilla *et al.* (2018) reported that the *ACSM5.P* was also the most significantly associated SNP with the *ACSM5* gene expression in adipose tissue in BC1_LD animals. The ACSM5 is an acyl-CoA synthase involved in a preliminary step of the fatty acid β-oxidation pathway. This enzyme catalyses the activation of fatty acids by CoA to produce an acyl-CoA, and is then introduced in the mitochondria through the CPT system.

In summary, the *ACSM5.P* polymorphism is affecting the expression of the gene but we cannot discard that there are other genetic factors that may regulate the *ACSM5* gene expression in muscle.

Expression-GWAS studies were also performed for each backcross independently and 26, 32 and 25 eQTLs were identified in the BC1_LD, BC1_DU and BC1_PI animals, respectively. The two *cis*-eQTL regions of *ACSM5* and *IGF2* genes were segregating in all three backcrosses, suggesting that the Iberian boars and the three founder maternal breeds have different allelic frequencies for the polymorphisms regulating in *cis* the expression of these genes. In addition, six *trans*-eQTL hotspots regions, two in each backcross, regulating the expression of several genes were detected (Figure 4.1.).

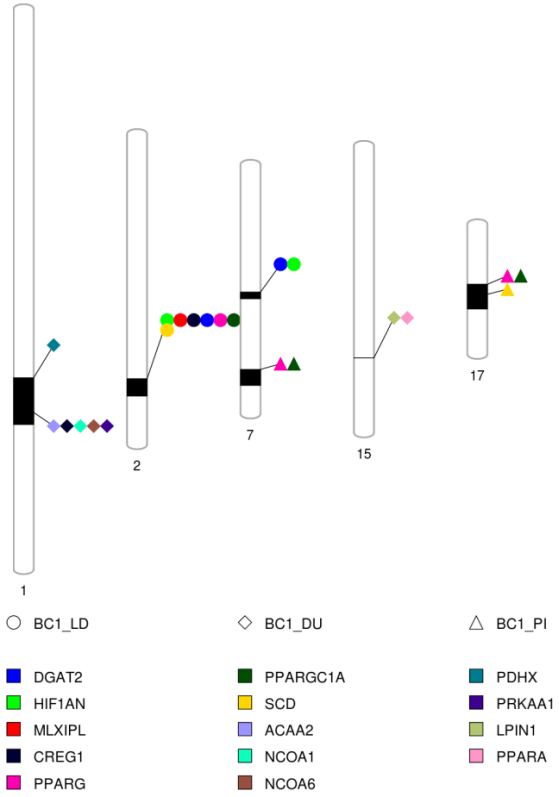


Figure 4.1. PhenoGram plot representing the six *trans*-eQTL hotspots regions found in the eGWAS individually. The shape indicates the backcross and the colour indicates the gene name as it is indicated in the legend.

Using the eQTL analysis, it is possible to identify potential transcription factors regulating the expression of several genes for a specific pathway (Sun, Yu and Li, 2007).

For instance, *NFKBIA* gene was a transcription factor described as a promising regulator to explain differences in gene expression variations of *HIF1AN* and *DGAT2* in BC1_LD animals, and *FOXA2* could act as a regulator of *PPARG*, *PPARGC1A* and *SCD* in the BC1_PI population. Moreover, other candidate genes were described as possible regulators of different genomic regions in the different populations, such as *TCF7* gene for the *CREG1* eQTL in the BC1_LD cross. In the BC1_DU population study, *PLIN2* gene was described as feasible regulator of *ACAA2*, *NCOA1*, *NCOA6* and *PDHX* genes and *AOX1* was suggested as a regulator of *LPIN1* and *PPARA* genes. In general, only three *trans*-eQTL regions were in common with the 3BCs study, which reinforces the presence of different regulatory mechanisms or allelic frequencies in each breed.

4.1.2. *IGF2*

In the late 1980s, a paternally expressed QTL for muscle growth and backfat thickness was identified in pig chromosome 2 (SSC2), containing the insulin-like growth factor 2 (*IGF2*) gene (Jeon *et al.*, 1999; Nezer *et al.*, 1999). Subsequently, the polymorphism *g.3072G>A* located in the intron 3 of the *IGF2* gene was described as the causal mutation for this QTL (Van Laere *et al.*, 2003). This polymorphism is widespread in different breeds, contributing to the porcine production and explaining the 15-30% of the phenotypic variation in muscle mass and 10-20% of the variation in backfat thickness (Jungerius *et al.*, 2004). Later, an association between *IGF2* gene expression and the percentage of intramuscular fat was reported (Aslan *et al.*, 2012). Moreover, an effect on both carcass and ham conformation, and an increase of MUFA content was found in animals carrying the *A* allele (López-Buesa *et al.*, 2013).

Previous studies of our group showed that the *g.3072G>A* polymorphism is associated with backfat thickness, carcass weight, ham weight and shoulder weight in a Large White commercial population and with backfat thickness, *longissimus* muscle area and ham weight traits in an Iberian x Landrace F2 cross (Estellé *et al.*, 2005). In addition, Puig-Oliveras *et al.* (2016) reported that the *IGF2:g.3072G>A* polymorphism was significantly associated with the *IGF2* mRNA variation in muscle of BC1_LD, although it was not the most significantly associated SNP, suggesting that other mutations may be the responsible of the *IGF2* mRNA variation (Puig-Oliveras *et al.*, 2016). In the current

thesis, we have genotyped the *IGF2g.3072G>A* polymorphism in the 3BCs animals and completed the *IGF2* expression analysis in animals of the BC1_DU and BC1_PI crosses. The eGWAS results showed that the *IGF2g.3072G>A* polymorphism was the most significantly associated SNP in the *IGF2 cis*-eQTL region with the muscle *IGF2* gene expression variation, explaining the 70% of the phenotypic variance. Then, we performed an eGWAS analysis for each backcross independently and we identified that the *IGF2:g.3072G>A* polymorphism was the most significantly associated SNP with the *IGF2* gene expression in muscle of BC1_DU and BC1_PI, explaining a high percentage of the phenotypic variance in both cases. Conversely, the *rs81322199* polymorphism located on SSC2 was the most significantly associated SNP in the BC1_LD animals (*p*-value=$1.45x10^{-15}$, *q*-value=$5.29x10^{-11}$) and explained around 42% of the phenotypic variance. In fact, the *IGF2:g.3072G>A* polymorphism was also significantly associated (*p*-value=$3.03x10^{-07}$, *q*-value=$7.87x10^{-04}$) but only accounts for the 22% of the phenotypic variance. This result may be explained by the low frequency of the *A* allele (0.2) in the BC1_LD population. Therefore, we cannot discard that another genetic variant in linkage disequilibrium with the *IGF2:g.3072G>A* SNP genotype is the causal variant of the *IGF2* gene expression differences observed in muscle of BC1_LD animals.

There was a lack of information about the role of *IGF2:g.3072G>A* polymorphism in adipose tissue and its fatty acid composition. Hence, we performed an expression analysis of *IGF2* gene in 355 animals belonging to BC1_LD, BC1_DU and BC1_PI populations in backfat adipose tissue. Subsequently, eGWAS analyses were carried out in two different ways: i) in a study with all 3BCs together and ii) in each backcross independently. Herein, the main objective of the eQTL analysis was the identification of chromosomic regions associated with the *IGF2* gene expression in backfat adipose tissue. The 3BCs study showed three chromosomal regions on SSC2 and SSC8 significantly associated with the *IGF2* mRNA variation in adipose tissue. In the SSC2 *cis*-eQTL region the *IGF2:g.3072G>A* polymorphism was the most significantly associated SNP with the *IGF2* mRNA variation. However, the SNP only explained a 25% of the phenotypic variance, suggesting that other genetic variants and/or environmental factors are involved in the regulation of the *IGF2* gene expression in adipose tissue. For example, an antisense transcript has been described as a regulator of the promoter 2,

3 and 4 transcription in post-natal muscle of animals carrying the A allele (Braunschweig *et al.*, 2004). In the SSC2 *trans*-eQTL region the *SF1* gene was annotated, and it was involved in gene expression regulatory mechanisms. On the contrary, in the SSC8 *trans*-eQTL no candidate genes were annotated, but we cannot discard this region due to the possible incomplete gene annotation.

Different results were obtained in the individually backcross eGWAS studies in comparison with the 3BCs study, with the exception of BC1_PI backcross where the *IGF2:g.3072G>A* polymorphism was the most significantly associated SNP for the *IGF2* mRNA variation. In BC1_LD, no eQTLs were identified, which could be explained by the same reason that was argued in muscle tissue, since the same animals were used. In BC1_DU, a polymorphism located on SSC8 was the most significantly associated SNP for the *IGF2* mRNA expression and the *IGF2:g.3072G>A* polymorphism was the most significantly associated SNP on SSC2, and explained a 24% of the phenotypic variance. Four *trans*-eQTLs regions were identified and strong candidate genes were mapped in two of them: *SF1* (SSC2) and *IGFBP1* and *IGFBP3* (SSC8).

In summary, in the 3BCs study, only the SSC2 eQTL region was found for the *IGF2* gene expression in muscle, being the *IGF2:g.3072G>A* polymorphism the most significantly associated SNP and explaining a high percentage of the phenotypic variance. Conversely, the *IGF2:g.3072G>A* polymorphism was also the most significantly associated SNP for the *IGF2* gene expression in adipose tissue, but had a lower effect and a *trans*-eQTL was found on SSC8.

Van Laere and collaborators (2003) demonstrated that the *IGF2* gene is imprinted in the muscle tissue and the methylation status of the region where the mutation is located abrogates the binding site for a transcription repressor called ZBED6 (Van Laere *et al.*, 2003; Huang *et al.*, 2014). In fact, a recent work demonstrated that the disruption of this ZBED6 binding site by CRISPR/Cas9 in porcine embryonic fibroblasts from an autochthonous Chinese pig breed, lead to an up-regulation of *IGF2* gene expression and myogenesis (Liu *et al.*, 2019).

In our study significant differences in gene expression between the *IGF2:g.3072G>A* heterozygous genotypes with different paternally-inherited alleles were found in

muscle but not in adipose tissue (Figure 4.2.), suggesting an imprinting effect in muscle, but not in adipose tissue. In order to investigate the imprinting status, differential allelic expression analysis between the two heterozygous genotypes in muscle and adipose tissue was performed.

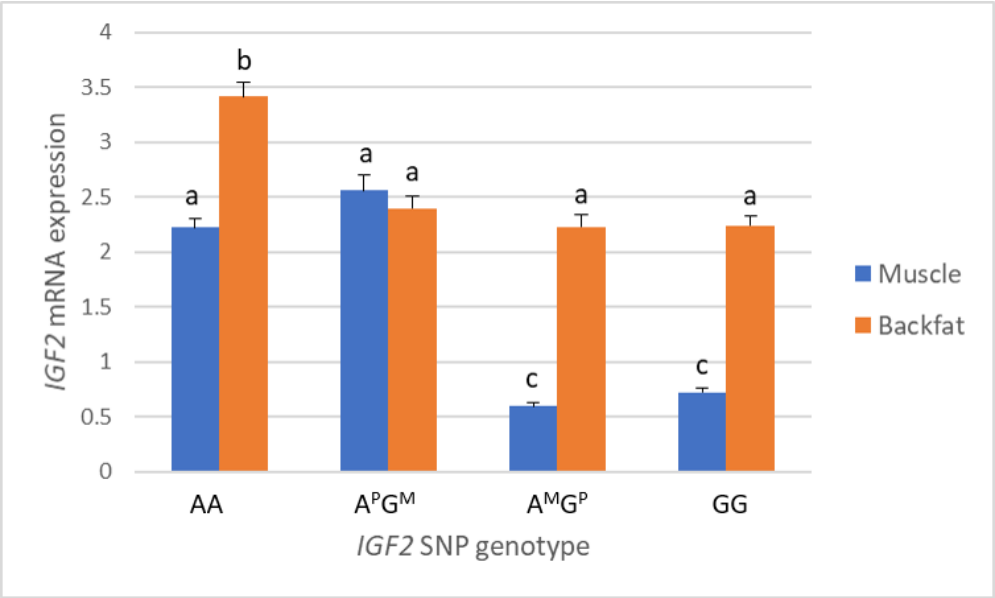


Figure 4.2. Plot of relative quantification of *IGF2* mRNA expression levels in muscle and adipose tissue, according to the *IGF2:g.3072G>A* genotype. Values with different superscript letters (a, b, and c) indicate significant differences between groups (*p*-value < 0.05). Paternally inherited allele was deduced and A^PG^M means a paternally inherited *A* allele and maternal inherited *G* allele, while A^MG^P represents a maternal inherited *A* allele and paternal inherited *G* allele.

The pyrosequencing method was chosen to carry out the detection of differential allelic expression because it allows the detection of differences as small as 4% (Neve *et al.*, 2002; Wasson *et al.*, 2002). The *IGF2:g.3072G>A* polymorphism was located in an intronic/promoter region, hence the pyrosequencing analysis was not possible, instead a polymorphism located in the 3'UTR region was used, which is in complete linkage disequilibrium with the *IGF2:g.3072G>A* polymorphism. Paternally inherited alleles were deduced from the genotypes of the progenitors and pedigree information, and seven animals carrying each haplotype were analysed. The results showed higher levels of the *A* allele percentage in animals inheriting the *A* allele from his father than

in animals receiving the *G* allele, in both muscle and adipose tissue. Therefore, there is imprinting of the *IGF2* gene in the two tissues. However, in muscle tissue higher *IGF2* mRNA levels were found in animals with the paternally-inherited *A* allele (A^P), while in adipose tissue the highest *IGF2* gene expression was observed in animals carrying the *AA* genotype, and no significant differences were observed between the A^PG^M and A^MG^P genotypes. These results can be explained by a higher expression of the *G* allele in adipose tissue, which can be due to a different methylation pattern and consequently a reduction of binding of the ZBED6 repressor. Hence, both *IGF2:g.3072G>A* SNP genotype and the imprinting model are explaining the *IGF2* gene expression differences observed in both muscle and adipose tissues.

We studied the association between the *IGF2:g.3072G>A* polymorphism and the fatty acid composition measured in adipose tissue. The *IGF2:g.3072G>A* polymorphism was the most significantly associated with linoleic (C18:2(n-6)), hexadecenoic (C16:1(n-9)), oleic (C18:1(n-9)), α-linoleic (C18:3(n-3)), arachidonic (C20:4(n-6)) fatty acids and the MUFA/PUFA ratio in the 3BCs study (Table 4.2.). In summary, the analysis of *IGF2* expression and fatty acid composition measured in adipose tissue revealed that homozygous *AA* animals, which showed the highest *IGF2* gene expression, presented a higher percentage of PUFA and a lower MUFA content in comparison with the other genotypes.

In contrast, in the analysis within each backcross the *IGF2:g.3072G>A* polymorphism was not always the most significantly associated SNP with fatty acid composition on SSC2. For instance, in BC1_LD the *IGF2:g.3072G>A* polymorphism was not significantly associated with fatty acid composition, probably due to the low *A*-allele frequency as explained above, and in BC1_DU no significant SNPs were found. Conversely, in BC1_PI animals the *IGF2:g.3072G>A* SNP was the most significantly associated one in four of the six fatty acid traits analysed (Table 4.2.).

Table 4.2. Summary of most significantly associated SNPs with fatty acid composition traits in the 3BCs, BC1_LD, BC1_DU and BC1_PI populations. NS means no significant associated SNPs were identified.

Trait	3BCs	BC1_LD	BC1_DU	BC1_PI
C16:1(n-9): hexadecanoic acid	*IGF2:g.3072G>A*, *p*-value=4.04x10^{-07}	*rs81322199*, *p*-value=9.63x10^{-07}	NS	NS
C18:2(n-6): linoleic acid	*IGF2:g.3072G>A*, *p*-value=6.44x10^{-09}	*rs81355859*, *p*-value=3.22x10^{-07}	NS	*IGF2:g.3072G>A*, *p*-value=1.79x10^{-05}
C18:1(n-9): oleic acid	*IGF2:g.3072G>A*, *p*-value=4.18x10^{-07}	*rs81287787*, *p*-value=4.39x10^{-06}	NS	*IGF2:g.3072G>A*, *p*-value=1.04x10^{-07}
C18:3(n-3): α-linoleic acid	*IGF2:g.3072G>A*, *p*-value=3.30x10^{-06}	*rs81316644*, *p*-value=8.04x10^{-06}	NS	*rs81312355*, *p*-value=3.80x10^{-05}
C20:4(n-6): arachidonic acid	*IGF2:g.3072G>A*, *p*-value=9.82x10^{-08}	NS	NS	*IGF2:g.3072G>A*, *p*-value=2.79x10^{-05}
MUFA/PUFA ratio	*IGF2:g.3072G>A*, *p*-value=2.51x10^{-09}	*rs81355859*, *p*-value=1.02x10^{-06}	NS	*IGF2:g.3072G>A*, *p*-value=1.67x10^{-07}

Adipose tissue is involved in lipogenesis, lipid storage and it is the main source of free fatty acids (O'Hea and Leveille, 1969; Kershaw and Flier, 2004; Nguyen *et al.*, 2008). It has a high content of PUFAs, such as the linoleic and α-linoleic essential fatty acids, which are involved in meat quality traits. On the contrary, MUFAs contributes to a high quality of cured products and improve meat flavour.

It has been reported an inverse relationship between the amount of α-linoleic measured in backfat and the backfat thickness (Wood *et al.*, 2008), although we could not find an association between the *IGF2:g.3072G>A* polymorphism and backfat thickness in our population. In the 3BCs study, the *rs81214179* SNP, which is located at 8.9 Mb of the *IGF2:g.3072G>A* SNP, was the most significantly associated SNP with the

backfat thickness trait (*p*-value=3.44x10^{-08}). At this position, three desaturases were mapped: *FADS1*, *FADS2* and *FADS3*. Desaturases are involved in the synthesis of highly unsaturated fatty acids from the essential fatty acids provided by the diet (Nakamura and Nara, 2004), and were proposed as strong candidate genes involved in fatty acid content.

In a future work it will be interesting to deep study the *FADS* gene family. For instance, its gene expression, which was reported to be higher in liver and adipose tissue than in muscle in pigs (Taniguchi *et al.*, 2015), as well as the possible linkage disequilibrium with the *IGF2* polymorphism. Moreover, to study polymorphisms located in the *FADS* genes and their association with the fatty acid composition. Nevertheless, we cannot discard that other genes located on SSC2 are also involved in the variability of fatty acid composition in adipose tissue.

4.1.3. miRNA 33 family

Transcription factors are well-known to play a key role in the regulation of the expression of genes involved in fatty acid metabolism. *Sterol regulatory element binding transcription factor* (*SREBF*) and *peroxisome proliferator activated receptor* (*PPAR*) are the two major transcription factors that have been reported to modulate and control the transcription of genes involved in fatty acid oxidation and ketogenesis pathways. They are engaged in the regulation of the expression of desaturases, such as *ELOVL6* and *SCD*, through the insulin signalling pathway (Guillou *et al.*, 2004; Nakamura and Nara, 2004). However, gene expression is not only regulated at the transcriptional level and miRNAs are emerging as important post-transcriptional regulators. miRNAs are predicted to regulate the expression of around 20-30% of the genes and hundreds of miRNAs have been identified in various farm animal species. Until now, some miRNAs have been involved in the regulation of lipid metabolism in different porcine tissues. For example, miR-210 and miR-27 were involved in adipogenesis and miR-374b and miR-130b in lipid metabolism in adipose tissue, whereas miR-122 plays a key role in cholesterol, fatty acid and lipid metabolism in liver. In muscle, several miRNAs have important functions in myogenesis and during

muscle development, and the frequently highest expressed miRNAs in this tissue are miR-1, miR-133 and miR-206 (Wang, Gu and Jiang, 2013; Song *et al.*, 2018). It has been reported in different species that the members of the miR-33 family: *mir-33a* and *mir-33b* are co-transcribed with the *SREBF2* and *SREBF1* transcription factors, respectively. *SREBF1* enhances the transcription of genes involved in fatty acid synthesis while *SREBF2* regulates genes involved in cholesterol synthesis and uptake (Shimano, 2001; Dávalos *et al.*, 2011).

In the present work, the expression of miR-33a and miR-33b in liver, adipose tissue and muscle was studied in 42 animals from the BC1_LD backcross, and its correlation with *in-silico* predicted target genes and fatty acid composition traits was calculated.

First, we did not find significant correlations between the miR-33b and the *SREBP1* gene expression in any tissue, while a significant positive correlation was found for miR-33a and *SREBF2* gene expression in liver, suggesting a different transcription pattern between miR-33a and miR-33b. Furthermore, in agreement with our results, Taniguchi *et al*. (2014) have reported a low correlation between miR-33b and *SREBF1* gene expression in pig adipose tissue (Taniguchi *et al.*, 2014).

In addition, different tissue-specific expression patterns for miR-33a and miR-33b suggested that tissue-specific mechanisms are regulating the expression of both miRNAs. On the other hand, high correlations between miR-33a and miR-33b in adipose tissue and muscle were found (Figure 4.3.), and considering that they have the same seed sequence, we think that both miRNAs may play similar roles in these tissues. In pig adipose tissue, miR-33b has been involved in lipogenesis and adipogenesis pathways (Taniguchi *et al.*, 2014). Nevertheless, further studies are needed to deepen the role of miR-33a and if both members of the miR-33 family have a similar function in adipose tissue. On the other hand, there is a lack of information about the role of these miRNAs in muscle tissue in the bibliography. Muscle is implicated in the regulation of lipid metabolism, being a key tissue for glucose uptake and storage, and an amino acid reservoir for protein synthesis or energy production (Meyer *et al.*, 2002). In liver, lower expression of miR-33a and miR-33b in comparison with other tissues was found, and the low correlation values obtained between both

miRNAs (Figure 4.3.) may suggest that each miR-33 family member has a different function in liver. These results are in accordance with other studies reporting that miR-33a and miR-33b are co-transcribed with their host genes, *SREBF2* and *SREBF1,* respectively, and play different functions: miR-33a has been described to participate in the regulation of cholesterol pathway genes, while miR-33b participates in fatty acid metabolism and insulin signalling (Gerin *et al.*, 2010; Horie *et al.*, 2010; Marquart *et al.*, 2010; Najafi-Shoushtari *et al.*, 2010; Rayner *et al.*, 2010; Dávalos *et al.*, 2011; Ramirez *et al.*, 2013).

As stated above, previous works of our group have studied the expression of candidate genes involved in lipid metabolism in liver, adipose tissue and muscle of BC1_LD animals (Puig-Oliveras *et al.*, 2016; Ballester *et al.*, 2017b; Revilla *et al.*, 2018). Target genes of miR-33a and miR-33b were selected from the previous list of candidate genes with gene expression data to perform correlation analysis among genes and miRNAs. A negative correlation between miR-33b and *CPT1A* expression level and positive correlations between miR-33a and *PPARGC1A* and *USF1* found in liver (Figure 4.3.) reinforced the hypothesis that, although both miRNAs have been implicated in fatty acid β-oxidation, miR-33b plays a more relevant role compared to miR-33a. However, the lack of many significant correlations between miRNAs and target genes indicates the complexity of the miRNA-mediated gene regulation, which depends on the miRNAs location, both miRNA and target mRNAs abundances, the affinity of the miRNA-mRNA interaction and the cell type or state among others (O'Brien *et al.*, 2018).

Finally, we have found positive correlations between miR-33a/b measured in liver and adipose tissue and SFAs or total SFA content and negative correlations with PUFAs or total PUFA content (Figure 4.3.). These results agree with previous studies of our group, where higher mRNA levels of genes involved in lipolysis and cholesterol homeostasis, and lower expression levels of lipogenic genes in liver and adipose tissue were found in pigs with a higher PUFA content (Ramayo-Caldas *et al.*, 2012; Corominas *et al.*, 2013a). Moreover, the correlation between liver miR-33a expression and adipose tissue fatty acid composition is in agreement with the cholesterol and lipogenesis pathways interaction (Tsai, Romsos and Leveille, 1975; Knight *et al.*, 2005). Hence, miR-33 family can be implicated in the determination of fatty acid metabolism.

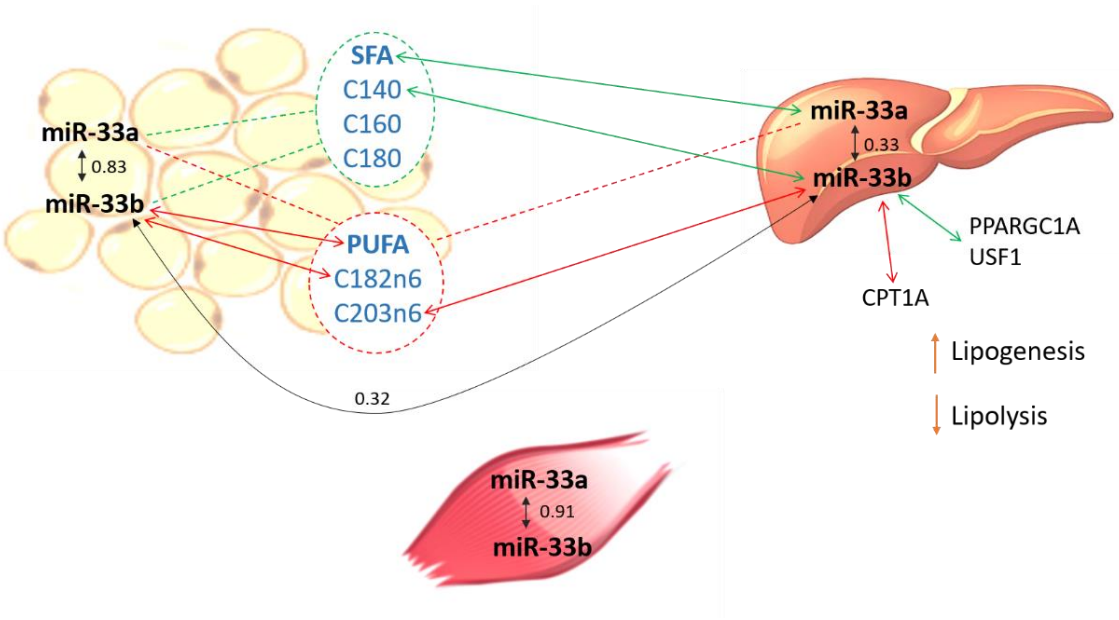


Figure 4.3. Schematic representation of all the correlation results. Only significant correlations were represented. Green lines mean positive correlations and red lines correspond to negative correlations. Discontinues lines indicates correlation with all the fatty acid composition traits circled.

4.1.4. *ELOVL6*

The *ELOVL6* gene is the responsible of the elongation of SFA and MUFAs with 12-16 carbons to C18, which clearly affects the content of palmitic (C16:0) and palmitoleic (C16:1n7) fatty acids (Moon *et al.*, 2001; Jakobsson, Westerberg and Jacobsson, 2006). Previous works of our group have analysed the *ELOVL6* gene as the main positional candidate gene explaining the phenotypic variation of a QTL on SSC8 for the percentage of palmitic and palmitoleic fatty acids content and the elongation ratios of C18:0/C16:0 and C18:1n-7/ C16:1n-7 in both intramuscular fat and backfat. The *ELOVL6:c.-533C>T* polymorphism has been associated with the palmitic and palmitoleic fatty acid content, as well as the elongation ratios in muscle and backfat in BC1_LD animals. Furthermore, the polymorphism was associated with the *ELOVL6* mRNA levels in adipose tissue (Corominas *et al.*, 2013b). Subsequently, the BAC screening and sequencing approach performed to study the 3'UTR of *ELOVL6* identified other

polymorphisms, and the *ELOVL6:c.-1922C>T* was associated with the palmitic and palmitoleic acid content in both intramuscular fat and backfat, although the *ELOVL6:c.-533C>T* SNP presented a higher association. Finally, for the functional characterization of the *ELOVL6* promoter, transcription factors binding sites were studied. Hence, Corominas *et al.* (2015) proposed that the differential occupancy of *ELOVL6* promoter by ERα*,* where the *ELOVL6:c.-394G>A* SNP is located, affects the methylation levels and, subsequently, the occupancy by *SREBF1* and *SP1* transcription factors, followed by a decrease in the gene expression (Corominas *et al.*, 2015).

At the beginning the ChIP technique was widely used to study the presence of specific histone modifications at some specific DNA regions using antibodies. Later, the analysis of this DNA by qPCR was emerged to determine the abundance of a region of interest, such as a transcription factor binding site. This technique is well stablished for many model systems but for other organisms is still challenging and time consuming (Haring *et al.*, 2007).

During my internship at INRA, France, as part of my thesis, a ChIP-qPCR protocol has been established and optimised in liver tissue samples with the objective to validate the occupancy of ERα, *SREBF1* and *SP1* transcription factors in the promoter region of the *ELOVL6* gene, were the *ELOVL6:c.-394G>A* mutation is located. Details about the ChIP-qPCR protocol are in Annex 7.4. A high range of Ct values was obtained and was not possible to differentiate between samples and controls and between animals carrying the *ELOVL6:c.-394G* allele and animals carrying the *ELOVL6:c.-394A* allele. Hence, this preliminary qPCR results suggest us that we should use another qPCR methodology, such as Taqman probes or digital PCR, because they should be more suitable to perform the ChIP-qPCR in our material.

4.2. Muscle transcriptome study using RNA-Sequencing

Muscle is the major component of the meat, together with adipose, epithelial, connective and nervous tissues. As already introduced, meat quality is a current topic and its known to be influenced by fatty acid composition and deposition of the muscle.

Both dietary fatty acids and genetics are determining the fatty acid composition of the muscle tissue (Wood *et al.*, 2008).

One of the major applications of RNA-Seq studies has been the evaluation of differentially expressed genes between groups, for example with extreme values for a specific phenotype. As mentioned before, previous works of our group studied the liver, adipose tissue and muscle transcriptomes in two groups of extreme animals for fatty acid composition in the muscle. These studies revealed a decrease in fatty acid oxidation in the liver, an increase in *de novo* lipogenesis in the adipose tissue and an increase in fatty acid and glucose uptake as well as an enhanced lipogenesis in the muscle of pigs with higher levels of MUFA and SFA content (Ramayo-Caldas *et al.*, 2012; Corominas *et al.*, 2013a; Puig-Oliveras *et al.*, 2014).

Considering the importance of the muscle tissue in the meat porcine industry, in this thesis we studied the *Longissimus dorsi* muscle transcriptome of 132 BC1_DU animals by RNA-Seq. To identify gene-expression regulators, eGWAS studies were performed and a total of 324 eQTL regions for the expression of 291 genes were identified. Most of the eQTLs were classified as *cis*-eQTLs, 247 regions, while only 77 regions showed a *trans*-effect. There is some controversy about the abundance of *cis* and *trans*-eQTLs. Studies using model organisms have been identified more *trans*-eQTLs and regulatory hotspots than studies performed in humans (Gilad, Rifkin and Pritchard, 2008; Cheung and Spielman, 2009). In addition, it was demonstrated that *trans*-eQTL are more difficult to identify in humans compared to experimental crosses, which have few alleles segregating at high frequencies and large linkage blocks which increase the statistical power (Albert and Kruglyak, 2015). Compared to *cis*-eQTLs, *trans*-eQTLs have been reported to be less frequent because they have lower effects and their analysis is computationally challenging, as well as being more difficult to replicate across studies (Gilad, Rifkin and Pritchard, 2008; Cheung and Spielman, 2009; Steibel *et al.*, 2011; Westra *et al.*, 2013). Although *trans*-eQTLs are less studied and their effects on gene expression variation and their biological mechanisms are still unclear, some studies in humans have reported that some *trans*-eQTL associations can be explained by a *cis* regulatory mechanism (Pierce *et al.*, 2014; Yao *et al.*, 2017). However, other studies showed more *trans* than *cis*-eQTLs (Morley *et al.*, 2004; Myers *et al.*, 2007; Ponsuksili

et al., 2008; Liaubet *et al.*, 2011; Cánovas *et al.*, 2012; Leal-Gutiérrez, Elzo and Mateescu, 2020). Some experimental factors, like sample size, affect the statistical power of eQTL mapping (Albert and Kruglyak, 2015) and may explain these discrepancies. In fact, it has been demonstrated that in large datasets, *trans*-eQTLs are more abundant than *cis*-eQTLs (Cheung *et al.*, 2010).

When a SNP located in a *trans*-eQTL affect the expression of several genes, the region is defined as *trans*-eQTL hotspot (Breitling *et al.*, 2008). In the current work, seven *trans*-eQTL hotspots were identified and several interesting transcription factors were mapped in some of these regions: *KLF2*, *SMARCA4*, and *TEAD1* (SSC2), *GDF6* (SSC4) and *ZFP36* (SSC6). These transcription factors regulate genes implicated in cell proliferation and differentiation signalling pathways of muscle tissue.

Moreover, a strong correction for multiple testing at local and global level was applied, which reduced the number of significant eQTLs detected. In the previous eGWAS performed for 45 lipid-related genes from the muscle of animals of the BC1_DU population, a list of eQTLs was provided, of which not all have been identified in the current study by RNA-Seq. This may be explained by the strong filtering of eQTL signals applied at global level, the different number of animals used in both experiments, and the technical differences between both methods (qPCR and RNA-Seq).

In the present work, the most significant associations were found for the *HGFAC* and the *HUS1* genes, which are members of crucial pathways in the muscle tissue: tissue regeneration and repair and response to DNA damage, respectively (O'Connell, Walworth and Carr, 2000; Fukushima *et al.*, 2018). In the functional studies of the 291 genes with eQTLs, different pathways were identified with the IPA program (Table 4.3.) and the metabolic process was the most relevant one reported using the WebGestalt program (Wang *et al.*, 2017).

Table 4.3. Top three ingenuity canonical pathways generated by IPA.

Canonical Pathways	*P*-value	Genes
Granzyme B Signaling	$7.01x10^{-04}$	*APAF1*, *DFFA*, *PRKDC*
Glutathione-mediated Detoxification	$5.44x10^{-03}$	*GSTA4*, *GSTM4*, *MGST2*
NRF2-mediated Oxidative Stress Response	$5.60x10^{-03}$	*EPHX1*, *FTL*, *GPX2*, *GSTA4*, *GSTM4*, *HERPUD1*, *MGST2*

Subsequently, we studied potential gene expression regulators of these genes. Transcription factors and miRNAs are involved in *trans*-acting gene regulation and can regulate several target genes by binding to their *cis*-regulatory elements. Most of the genes are controlled by a combination of *trans*-acting factors. Furthermore, transcription factors and miRNAs play a key role in the regulation at transcriptional and post-transcriptional level, respectively (Hobert, 2008). Therefore, in the current work different networks were generated for the most important transcription factors, miRNAs and nuclear receptors found in the muscle transcriptome data.

The most relevant gene expression regulators identified in the pig muscle tissue were *HNF4A*, *KLF3*, *E3F4*, *mir-483* and *RORC*. Whereas the *HNF4A* transcription factor is mainly expressed in liver and adipose tissue, in the current work was identified as a regulator of the muscle gene expression. In liver, the *HNF4A* gene plays a key role in the control of the expression of genes involved in lipid metabolism as well as to maintain normal lipid homeostasis (Yin *et al.*, 2011). In adipose tissue, the *HNF4A* gene has been identified as an upstream regulator activated in Iberian pigs and has been related with an inflammatory response (Benítez *et al.*, 2019). The *KLF3* is a broadly transcription factor involved in the regulation of genes for the different kind of muscles (Himeda *et al.*, 2010) and the *E2F4* transcription factor has been described to play a role in fatty acid transport and degradation, as well as lipid synthesis in bovines (Zhao *et al.*, 2016). The *mir-483* has been co-transcribed with the *IGF2* gene, which may indicates the possible role of the miRNA in the regulation of metabolism pathways (Ma *et al.*, 2011). Finally, the *RORC* nuclear receptor forms heterodimers with *PPARG*

and *RXR* and interacts with genes involved in the synthesis of cholesterol, steroids and other lipids, glutathione metabolism and retinoic acid pathways (Figure 4.4.).

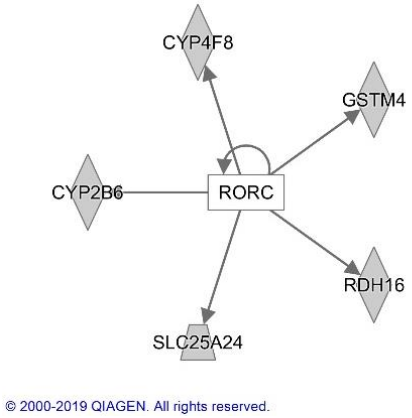


Figure 4.4. Network generated by IPA of *RORC* gene and their target genes.

These transcription factors, miRNAs and nuclear receptors are potential candidates to regulate muscle gene expression and should be further analysed, as well as their target genes and pathways, to better understand the main gene expression regulatory mechanisms in the pig muscle.

4.3. Future perspectives and challenges

The continuous development of high-throughput sequencing technologies and the decrease in sequencing costs will allow in the coming years to increase the number of complete genome sequencing studies of selected animals in pigs, including different breeds, animals with extreme phenotypes for important traits, etc. In addition, the transcriptome will be analysed in a greater number of animals and in different tissues per animal. However, it will be necessary to filter, select and validate the SNPs found through NGS tools in order to determine the role of these polymorphisms on the QTLs to perform gene-assisted selection for the phenotypic traits without influencing other phenotypes. Previous association analysis and RNA-Seq studies of our group were performed with the prior porcine assembly *Sus Scrofa 10.2,* limiting the results obtained due to the incomplete genome annotation. Hence, analysis with the current porcine assembly may provide new results of unannotated genes or miRNAs that have been reported to play relevant functions in metabolic pathways affecting lipid

metabolism. Nevertheless, there is still limited annotation information about miRNAs and other non-coding RNAS, which have been reported to play an important role in porcine gene expression regulation. Some of the candidate gene variants are already used as markers by the porcine industry (Andersson and Georges 2004) although most of the genes and variants behind complex traits are still unknown.

On the other hand, recently there has been an outstanding interest in high-throughput omics technologies, including genomics, transcriptomics, epigenomics, proteomics, metagenomics and metabolomics among others and their application in animal production (MacKay, Stone and Ayroles, 2009). The integration of the data obtained from all these "omics" and animal phenotypes using systems biology approaches is necessary to deepen the analysis of complex traits. In line with this, a new research line of our group has included the study of microbiome and its interaction with the growth and fatness of pigs. Moreover, exploring the metabolomic data may help in the understanding of how different metabolic pathways can affect the gene expression variance, or proteomic studies can contribute to the study of diseases and the use of proteins as a target or biomarkers. Finally, epigenomics will be an interesting layer to study because its implication in gene expression regulation.

In the present work, some new candidate genes of eQTLs were identified, and gene expression analyses were performed for functional validation. However, further analyses are needed to identify or functionally validate the potential causal mutations of the eQTLs described. Functional studies of genetics variants involved in the variation of the phenotypic traits are still scarce although they are essential for the correct biological interpretation of the results. For example, ChIP-seq or ChiP-qPCR may be useful to validate gene interactions and site-directed mutagenesis to study causal mutations. In addition, genome editing techniques are suitable to modify at the same time multiple *loci* related with a specific trait in multiple animals, in an efficient way. For instance, genome editing has been used in livestock animals to increase disease resistance, to help in the adaptation to farm or environmental conditions, to improve fertility and growth, and to ensure animal welfare (Ricroch 2019). For example, the well-known CRISPR-Cas9 system has been tested to validate the possible causal mutations of the QTLs for production traits (Yang and Wu, 2018).

Regarding the muscle gene expression study of 45 lipid-related genes, a mutation located in promoter region of the *ACSM5* gene was validated as a potential variation to explain differences in the muscle gene expression in different genetic backgrounds, although further functional validations should be done to confirm if the polymorphism is the causal mutation. In addition, a deep study of the transcription factors mapped in the breed-specific hotspot eQTL regions would be important to identify gene expression regulators of porcine lipid metabolism.

Moreover, the *IGF2* polymorphism showed an important function not only in muscle but also in adipose tissue, where we suggested that can play an important role in fatty acid composition. However, further studies may be done, for example analysing the *FADS* genes and their implication in the determination of fatty acid composition. In addition, epigenetic studies such as methylation studies may be necessary to better understand the adipose tissue *IGF2* gene expression regulation.

For the *ELOVL6* gene, an improvement of the ChIP-qPCR technique is necessary to confirm the binding of the three different transcription factors involved in the gene expression pattern and consequently in the fatty acid composition.

The *mir-33a* and *mir-33b* were evaluated as potential candidate genes to determine fatty acid composition due to its location on *SREBF* family transcription factors and their large number of target genes involved in lipid metabolism. Additional studies should contemplate a higher number of animals and the functional validation of the interaction between the miRNAs and some of their target genes, as well as the quantification of the amount of protein of some target genes.

Finally, in the preliminary results of the muscle RNA-Seq study in 132 BC1_DU animals, some genes and pathways were described to be involved in the muscle gene expression regulation, but further parameters should be tested in the eGWAS analyses and the eQTL annotation. Later, a detailed study of the proposed candidate regulators of gene expression should be performed. At last, other analysis can be done, such as identifying genes differentially expressed according to phenotypes for meat quality traits.

Conclusions

Chapter 5

1. The gene expression pattern of 45 candidate genes for fatty acid composition in muscle was studied by multiple real-time qPCR in microfluidic arrays in 355 pigs from three different genetic backgrounds. The eGWAS identified a total of ten *trans*-acting eQTLs for *ACSM5*, *ACSS2*, *ATF3*, *DGAT2*, *FOS,* and *IGF2* genes and two *cis*-acting eQTLs for *ACSM5* and *IGF2* genes.

2. The within-backcross eGWAS revealed different eQTL regions for each backcross, suggesting that breed-specific genetic variants are regulating the expression of candidate genes. Six *trans*-eQTL hotspots, two per backcross, regulating the expression of up to seven candidate genes were identified.

3. Polymorphisms located in the promoter regions of the *IGF2* (*IGF2:g.3072G>A*) and *ACSM5* (*rs331702081*) genes are explaining a high percentage of the expression variation of these genes in the pig muscle.

4. The *IGF2:g.3072G>A* SNP is the most significantly associated polymorphism with the backfat adipose tissue *IGF2* gene expression in an eGWAS study with 355 animals belonging to different genetic backgrounds. In the within-backcross analysis, *IGF2:g.3072G>A* is also the most significant SNP of SSC2 in Duroc and Pietrain backcrosses, but not in Landrace where the polymorphism is segregating at lower frequency.

5. The *IGF2* gene is imprinted in both muscle and adipose tissue, being higher the expression of the paternally-inherited A allele than the maternally-inherited allele. However, stronger difference between the paternally and maternally inherited *A* allele expression was observed in muscle.

6. The *IGF2:g.3072G>A* polymorphism has been associated with the MUFA/PUFA ratio and the oleic, hexadecenoic, linoleic, α-linoleic, and arachidonic fatty acids measured in backfat, and animals carrying the *A* allele showed a higher PUFA and lower MUFA content.

7. The expression of miR-33a and miR-33b was studied in liver, adipose tissue and muscle of 42 Iberian x Landrace crossbreed pigs, showing different gene expression regulatory mechanisms among tissues. High expression correlations between miR-33a and miR-33b were observed in muscle and adipose tissue, but not in liver, suggesting that both miRNAs are differentially regulated and have distinct functions in liver.

8. Significant correlations were observed among adipose tissue fatty acid composition and the liver and adipose tissue miR-33a and miR-33b expressions, reinforcing the role of these miRNAs in the regulation of lipid metabolism.

9. The muscle transcriptome of 132 Iberian x Duroc crossbreed pigs has been analysed by RNA-Seq. The eGWAS allowed the identification 247 *cis*-eQTL and 77 *trans*-eQTL regions for the expression levels of 291 genes, which are mainly involved in metabolic pathways.

10. The functional analysis identified *HNF4A*, *KLF3* and *E3F4* transcription factors, miRNA-438, and *RORC* nuclear receptor as the main regulators of the porcine muscle expression of 291 genes with significant eQTLs.

References

Chapter 6

Albert, F. W. and Kruglyak, L. (2015) 'The role of regulatory variation in complex traits and disease', *Nature Reviews Genetics*, 16(4), pp. 197–212.

Ameer, F., Scandiuzzi, L., Hasnain, S., Kalbacher, H. and Zaidi, N. (2014) 'De novo lipogenesis in health and disease', *Metabolism: Clinical and Experimental*, 63(7), pp. 895–902.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., et al. (1994) 'Genetic mapping of quantitative trait loci for growth and fatness in pigs', *Science*, 263(5154), pp. 1771–1774.

Archibald, A. L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. A. M., Harlizius, B., et al. (2010) 'Pig genome sequence - analysis and publication strategy', *BMC Genomics*, 11, p. 438.

Aslan, O., Hamill, R. M., Davey, G., McBryan, J., Mullen, A. M., Gispert, M., et al. (2012) 'Variation in the IGF2 gene promoter region is associated with intramuscular fat content in porcine skeletal muscle', *Molecular Biology Reports*, 39(4), pp. 4101–4110.

Ayuso, M., Fernández, A., Núñez, Y., Benitez, R., Isabel, B., Barragán, C., et al. (2015) 'Comparative analysis of muscle transcriptome between pig genotypes identifies genes and regulatory mechanisms associated to growth, Fatness and metabolism', *PLoS ONE*, 10(12), p. e0145162.

Ballester, M., Puig-Oliveras, A., Castelló, A., Revilla, M., Fernández, A. I. and Folch, J. M. (2017a) 'Association of genetic variants and expression levels of porcine FABP4 and FABP5 genes', *Animal Genetics*, 48(6), pp. 660–668.

Ballester, M., Ramayo-Caldas, Y., Revilla, M., Corominas, J., Castelló, A., Estellé, J., et al. (2017b) 'Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs', *Scientific Reports*, 7, p. 46539.

Ballester, M., Revilla, M., Puig-Oliveras, A., Marchesi, J. A. P., Castelló, A., Corominas, J., et al. (2016) 'Analysis of the porcine APOA2 gene expression in liver, polymorphism identification and association with fatty acid composition traits', *Animal Genetics*, 47(5), pp. 552–559.

Benítez, R., Trakooljul, N., Núñez, Y., Isabel, B., Murani, E., De Mercado, E., et al. (2019) 'Breed, diet, and interaction effects on adipose tissue transcriptome in iberian and duroc pigs fed different energy sources', *Genes*, 10(589), pp. 1–24.

Bergen, W. G. and Burnett, D. D. (2013) 'Topics in Transcriptional Control of Lipid Metabolism: from Transcription Factors to Gene-Promoter Polymorphisms',

Journal of Genomics, 1, pp. 13–21.

Bergen, W. G. and Mersmann, H. J. (2005) 'Comparative Aspects of Lipid Metabolism: Impact on Contemporary Research and Use of Animal Models', *The Journal of Nutrition*, 135(11), pp. 2499–2502.

Braunschweig, M. H., Van Laere, A. S., Buys, N., Andersson, L. and Andersson, G. (2004) 'IGF2 antisense transcript expression in porcine postnatal muscle is affected by a quantitative trait nucleotide in intron 3', *Genomics*, 84(6), pp. 1021–1029.

Breitling, R., Li, Y., Tesson, B. M., Fu, J., Wu, C., Wiltshire, T., et al. (2008) 'Genetical Genomics : Spotlight on QTL Hotspots', *PLoS Genetics*, 4(10), p. e1000232.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009) 'The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments', *Clinical Chemistry*, 55(4), pp. 611–622.

Bustin, S. A. and Nolan, T. (2004) 'Pitfalls of quantitative real- time reverse-transcription polymerase chain reaction', *Journal of Biomolecular Techniques*, 15(3), pp. 155–166.

Cai, Y., Yu, X., Hu, S. and Yu, J. (2009) 'A Brief Review on the Mechanisms of miRNA Regulation', *Genomics, Proteomics and Bioinformatics*, 7(4), pp. 147–154.

Cameron, N. D. (1990) 'Genetic and phenotypic parameters for carcass traits, meat and eating quality traits in pigs', *Livestock Production Science*, 26(2), pp. 119–135.

Cánovas, A., Pena, R. N., Gallardo, D., Ramírez, O., Amills, M. and Quintanilla, R. (2012) 'Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population', *PLoS ONE*, 7(4), p. e35583.

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

Cao, H., Gerhold, K., Mayers, J. R., Wiest, M. M., Watkins, S. M. and Hotamisligil, G. S. (2008) 'Identification of a Lipokine, a Lipid Hormone Linking Adipose Tissue to Systemic Metabolism', *Cell*, 134(6), pp. 933–944.

Cardoso, T. F., Quintanilla, R., Castelló, A., González-Prendes, R., Amills, M. and Cánovas, Á. (2018) 'Differential expression of mRNA isoforms in the skeletal muscle of pigs with distinct growth and fatness profiles', *BMC Genomics*, 19(1), pp. 1–12.

Cheung, V. G., Nayak, R. R., Wang, I. X., Elwyn, S., Cousins, S. M., Morley, M., et al. (2010) 'Polymorphic cis- and trans-regulation of human gene expression', *PLoS*

Biology, 8(9), p. e1000480.

Cheung, V. G. and Spielman, R. S. (2009) 'Genetics of human gene expression: Mapping DNA variants that influence gene expression', *Nature Reviews Genetics*, 10(9), pp. 595–604.

Cirera, S., Birck, M., Busk, P. K. and Fredholm, M. (2010) 'Expression profiles of miRNA-122 and its target CAT1 in minipigs (Sus scrofa) fed a high-cholesterol diet', *Comparative Medicine*, 60(2), pp. 136–141.

Clop, A., Cercós, A., Tomàs, A., Pérez-Enciso, M., Varona, L., Noguera, J. L., et al. (2002) 'Assignment of the 2, 4-dienoyl-CoA reductase (DECR) gene to porcine chromosome 4', *Animal Genetics*, 33(2), pp. 164–165.

Clop, A., Ovilo, C., Perez-Enciso, M., Cercos, A., Tomas, A., Fernandez, A., et al. (2003) 'Detection of QTL affecting fatty acid composition in the pig', *Mammalian Genome*, 14(9), pp. 650–656.

Corominas, J., Marchesi, J. A., Puig-Oliveras, A., Revilla, M., Estellé, J., Alves, E., et al. (2015) 'Epigenetic regulation of the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs', *Genetics Selection Evolution*, 47(1), p. 20.

Corominas, J., Ramayo-Caldas, Y., Castelló, A., Muñoz, M., Ibáñez-Escriche, N., Folch, J. M., et al. (2012) 'Evaluation of the porcine ACSL4 gene as a candidate gene for meat quality traits in pigs', *Animal Genetics*, 43(6), pp. 714–720.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., et al. (2013a) 'Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition', *BMC Genomics*, 14, p. 843.

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

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Benítez, R., Fernández, A. I., et al. (2019) 'Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition', *PLoS ONE*, 14(8), p. e0220708.

Cui, J. X., Zeng, Y. Q., Wang, H., Chen, W., Du, J. F., Chen, Q. M., et al. (2011) 'The effects of DGAT1 and DGAT2 mRNA expression on fat deposition in fatty and lean breeds of pig', *Livestock Science*, 140(1–3), pp. 292–296.

D'Andrea, M., Dal Monego, S., Pallavicini, A., Modonut, M., Dreos, R., Stefanon, B., et

al. (2011) 'Muscle transcriptome profiling in divergent phenotype swine breeds during growth using microarray and RT-PCR tools', *Animal Genetics*, 42(5), pp. 501–509.

Damon, M., Wyszynska-Koko, J., Vincent, A., Hérault, F. and Lebret, B. (2012) 'Comparison of muscle transcriptome between pigs with divergent meat quality phenotypes identifies genes related to muscle metabolism and structure', *PLoS ONE*, 7(3), p. e33763.

Dávalos, A., Goedeke, L., Smibert, P., Ramírez, C. M., Warrier, N. P. and Andreo, U. (2011) 'miR-33a/b contribute to the regulation of fatty acid metabolism and insulin signaling', *PNAS*, 108(22), pp. 9232–9237.

Eaton, S., Bartlett, K. and Pourfarzam, M. (1996) 'Mammalian mitochondrial beta-oxidation', *Biochemical Journal*, 320, pp. 345–357.

Ernst, C. W. and Steibel, J. P. (2013) 'Molecular advances in QTL discovery and application in pig breeding', *Trends in Genetics*, 29(4), pp. 215–224.

Estellé, J., Fernández, A. I., Pérez-Enciso, M., Fernández, A., Rodríguez, C., Sánchez, A., et al. (2009a) 'A non-synonymous mutation in a conserved site of the MTTP gene is strongly associated with protein activity and fatty acid profile in pigs', *Animal Genetics*, 40(6), pp. 813–820.

Estellé, J., Mercadé, A., Noguera, J. L., Pérez-Enciso, M., Óvilo, C., Sánchez, A., et al. (2005) 'Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross', *Journal of Animal Science*, 83(12), pp. 2723–2728.

Estellé, J., Mercadé, A., Pérez-Enciso, M., Pena, R. N., Silió, L., Sánchez, A., et al. (2009b) 'Evaluation of FABP2 as candidate gene for a fatty acid composition QTL in porcine chromosome 8', *Journal of Animal Breeding and Genetics*, 126(1), pp. 52–58.

Estellé, J., Pérez-Enciso, M., Mercadé, A., Varona, L., Alves, E., Sánchez, A., et al. (2006) 'Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross', *Animal Genetics*, 37(6), pp. 589–591.

Esteve-Codina, A., Kofler, R., Palmieri, N., Bussotti, G., Notredame, C. and Pérez-Enciso, M. (2011) 'Exploring the gonad transcriptome of two extreme male pigs with RNA-seq', *BMC Genomics*, 12, p. 552.

Fernández, A. I., Pérez-Montarelo, D., Barragán, C., Ramayo-Caldas, Y., Ibáñez-Escriche, N., Castelló, A., et al. (2012) 'Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip', *BMC Genetics*,

13, p. 41.

Filipowicz, W., Bhattacharyya, S. N. and Sonenberg, N. (2008) ‘Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight?’, *Nature Reviews Genetics*, 9(2), pp. 102–114.

Freeman, T. C., Ivens, A., Baillie, J. K., Beraldi, D., Barnett, M. W., Dorward, D., et al. (2012) ‘A gene expression atlas of the domestic pig’, *BMC Biology*, 10, p. 90.

Frühbeck, G., Méndez-Giménez, L., Fernández-Formoso, J. A., Fernández, S. and Rodríguez, A. (2014) 'Regulation of adipocyte lipolysis', *Nutrition Research Reviews*, 27, pp. 63–93.

Fukushima, T., Uchiyama, S., Tanaka, H. and Kataoka, H. (2018) ‘Hepatocyte growth factor activator: A proteinase linking tissue injury with repair’, International Journal of Molecular Sciences, 19(11), pp. 1–11.

Gao, P., Cheng, Z., Li, M., Zhang, N., Le, B., Zhang, W., et al. (2019) ‘Selection of candidate genes affecting meat quality and preliminary exploration of related molecular mechanisms in the Mashen pig’, *Asian-Australasian Journal of Animal Sciences*, 32(8), pp. 1084–1094.

Gerin, I., Clerbaux, L. A., Haumont, O., Lanthier, N., Das, A. K., Burant, C. F., et al. (2010) ‘Expression of miR-33 from an SREBP2 intron inhibits cholesterol export and fatty acid oxidation’, *Journal of Biological Chemistry*, 285(44), pp. 33652–33661.

Gerrits, A., Li, Y., Tesson, B. M., Bystrykh, L. V, Weersing, E., Ausema, A., et al. (2009) ‘Expression Quantitative Trait Loci Are Highly Sensitive to Cellular Differentiation State’, *PLoS Genetics*, 5(10), p. e1000692.

Ghosh, M., Sodhi, S. S., Song, K. D., Kim, J. H., Mongre, R. K., Sharma, N., et al. (2015) ‘Evaluation of body growth and immunity-related differentially expressed genes through deep RNA sequencing in the piglets of Jeju native pig and Berkshire’, *Animal Genetics*, 46(3), pp. 255–264.

Gilad, Y., Rifkin, S. A. and Pritchard, J. K. (2008) ‘Revealing the architecture of gene regulation: the promise of eQTL studies’, *Trends in Genetics*, 24(8), pp. 408–415.

Goddard, M. E. and Hayes, B. J. (2009) ‘Mapping genes for complex traits in domestic animals and their use in breeding programmes’, *Nature Reviews Genetics*. Nature Publishing Group, 10(6), pp. 381–391.

Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., et al. (2018) ‘Meat consumption, health, and the environment’, *Science*, 361, p. 243.

Gol, S., Pena, R. N., Rothschild, M. F., Tor, M. and Estany, J. (2018) 'A polymorphism in the fatty acid desaturase-2 gene is associated with the arachidonic acid metabolism in pigs', *Scientific Reports*, 8(1), pp. 1–9.

González-Prendes, R., Quintanilla, R., Cánovas, A., Manunza, A., Cardoso, T. F., Jordana, J., et al. (2017) 'Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits', *Scientific Reports*, 7, pp. 1–9.

González-Prendes, R., Quintanilla, R., Mármol-Sánchez, E., Pena, R. N., Ballester, M., Cardoso, T. F., et al. (2019) 'Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine gluteus medius and longissimus dorsi muscles', *BMC Genomics*, 20, p. 170.

Gorni, C., Garino, C., Iacuaniello, S., Castiglioni, B., Stella, A., Restelli, G. L., et al. (2011) 'Transcriptome analysis to identify differential gene expression affecting meat quality in heavy Italian pigs', *Animal Genetics*, 42(2), pp. 161–171.

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012) 'Analyses of pig genomes provide insight into porcine demography and evolution', *Nature*, 491(7424), pp. 393–398.

Guillou, H., D'Andrea, S., Rioux, V., Jan, S. and Legrand, P. (2004) 'The surprising diversity of Δ6-desaturase substrates', *Biochemical Society Transactions*, 32(1), pp. 86–87.

Haley, C.S., Archibald, A., Andersson, L., Bosma, A.A., Davies, W., Fredholm, M., et al. (1990) 'The pig gene mapping project-Pigmap'. *Proceedings of the 4th World Congress on Genetics applied to livestock production*, pp. 67–70.

Hamill, R. M., Aslan, O., Mullen, A. M., O'Doherty, J. V., McBryan, J., Morris, D. G., et al. (2013) 'Transcriptome analysis of porcine M. semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction', *BMC Genomics*, 14, p. 453.

Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C. and Stam, M. (2007) 'Chromatin immunoprecipitation: Optimization, quantitative analysis and data normalization', *Plant Methods*, 3(1), pp. 1–16.

Hausman, G. J., Dodson, M. V., Ajuwon, K., Azain, M., Barnes, K. M., Guan, L. L., et al. (2009) 'Board-invited review: The biology and regulation of preadipocytes and adipocytes in meat animals', *Journal of Animal Science*, 87(4), pp. 1218–1246.

He, L. and Hannon, G. J. (2004) 'MicroRNAs: Small RNAs with a big role in gene regulation', *Nature Reviews Genetics*, 5(7), pp. 522–531.

Heidt, H., Cinar, M. U., Uddin, M. J., Looft, C., Jüngst, H., Tesfaye, D., et al. (2013) 'A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population', *Mammalian Genome*, 24(9–10), pp. 416–426.

Himeda, C. L., Ranish, J. A., Pearson, R. C. M., Crossley, M. and Hauschka, S. D. (2010) 'KLF3 Regulates Muscle-Specific Gene Expression and Synergizes with Serum Response Factor on KLF Binding Sites', *Molecular and Cellular Biology*, 30(14), pp. 3430–3443.

Hobert, O. (2008) 'Gene regulation by transcription factors and MicroRNAs', *Science*, 319(5871), pp. 1785–1786.

Horie, T., Ono, K., Horiguchi, M., Nishi, H., Nakamura, T., Nagao, K., et al. (2010) 'MicroRNA-33 encoded by an intron of sterol regulatory element-binding protein 2 (Srebp2) regulates HDL in vivo', *PNAS*, 107(40), pp. 17321–17326.

Houten, S. M., Violante, S., Ventura, F. V. and Wanders, R. J. A. (2016) 'The Biochemistry and Physiology of Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders', *Annual Review of Physiology*, 78(1), pp. 23–44.

Hu, Z. L., Dracheva, S., Jang, W., Maglott, D., Bastiaansen, J., Rothschild, M. F., et al. (2005) 'A QTL resource and comparison tool for pigs: PigQTLDB', *Mammalian Genome*, 16(10), pp. 792–800.

Huang, M., Chen, L., Shen, Y., Chen, J., Guo, X. and Xu, N. (2019) 'Integrated mRNA and miRNA profile expression in livers of Jinhua and Landrace pigs', *Asian-Australasian Journal of Animal Sciences*, 32(10), pp. 1483–1490.

Huang, Y.-Z., Zhang, L.-Z., Lai, X.-S., Li, M.-X., Sun, Y.-J., Li, C.-J., et al. (2014) 'Transcription factor ZBED6 mediates IGF2 gene expression by regulating promoter activity and DNA methylation in myoblasts.', *Scientific Reports*, 4, p. 4570.

Hulsegge, B., Calus, M., … A. H.-B.-B. of A. of and 2018, undefined (2019) 'Impact of merging different commercial lines on genetic diversity of the Dutch Landrace pigs'. *BioMed Central*, pp. 1–12.

Humphray, S. J., Scott, C. E., Clark, R., Marron, B., Bender, C., Camm, N., et al. (2007) 'A high utility integrated map of the pig genome', *Genome Biology*, 8(7), pp. R139.

Jakobsson, A., Westerberg, R. and Jacobsson, A. (2006) 'Fatty acid elongases in mammals : Their regulation and roles in metabolism', *Progress in Lipid Research*, 45, pp. 237–249.

Jansen, R. C. and Nap, J. P. (2001) 'Genetical genomics: The added value from

segregation', *Trends in Genetics*, 17(7), pp. 388–391.

Jeon, J. T., Carlborg, O., Törnsten, a, Giuffra, E., Amarger, V., Chardon, P., et al. (1999) 'A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus.', *Nature Genetics*, 21(2), pp. 157–158.

Jiang, S., Wei, H., Song, T., Yang, Y., Peng, J. and Jiang, S. (2013) 'Transcriptome Comparison between Porcine Subcutaneous and Intramuscular Stromal Vascular Cells during Adipogenic Differentiation', *PLoS ONE*, 8(10), p. e77094

Jungerius, B. J., Van Laere, A. S., Te Pas, M. F. W., van Oost, B., Andersson, L. and Groenen, M. (2004) 'The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan x European white pig intercross.', *Genetical research*, 84(2), pp. 95–101.

Kershaw, E. E. and Flier, J. S. (2004) 'Adipose tissue as an endocrine organ', *Journal of Clinical Endocrinology and Metabolism*, 89(6), pp. 2548–2556.

Kim, J. H., Lim, H. T., Park, E. W., Rodríguez, C., Silio, L., Varona, L., et al. (2006) 'Polymorphisms in the promoter region of the porcine acyl-coA dehydrogenase, medium-chain (ACADM) gene have no effect on fat deposition traits in a pig Iberian x Landrace cross', *Animal Genetics*, 37(4), pp. 430–431.

Kloosterman, W. P. and Plasterk, R. H. A. (2006) 'The Diverse Functions of MicroRNAs in Animal Development and Disease', *Developmental Cell*, 11(4), pp. 441–450.

Knight, B. L., Hebbach, A., Hauton, D., Brown, A. M., Wiggins, D., Patel, D. D., et al. (2005) 'A role for PPARα in the control of SREBP activity and lipid synthesis in the liver', *Biochemical Journal*, 389(2), pp. 413–421.

Korte, A. and Farlow, A. (2013) 'The advantages and limitations of trait analysis with GWAS: A review', *Plant Methods*, 9(1), p. 1.

Kouba, M. and Sellier, P. (2011) 'A review of the factors influencing the development of intermuscular adipose tissue in the growing pig', *Meat Science*, 88(2), pp. 213–220.

Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) 'MiRBase: From microRNA sequences to function', *Nucleic Acids Research*, 47(D1), pp. D155–D162.

Kuang, J., Yan, X., Genders, A. J., Granata, C. and Bishop, D. J. (2018) 'An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research', *PLoS ONE*, 13(5), p. e0196438.

L.van Dijk, E., Hélène, A., Jaszczyszyn, Y. and Thermes, C. (2014) 'Ten years of next-generation sequencing technology', *Trends in Genetics*, 30(9), pp. 418–426.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) 'Identification of tissue-specific MicroRNAs from mouse', *Current Biology*, 12(9), pp. 735–739.

Leal-Gutiérrez, J. D., Elzo, M. A. and Mateescu, R. G. (2020) 'Identification of eQTLs and sQTLs associated with meat quality in beef', *BMC Genomics*, 21, p. 104.

Lee, R., Feinbaum, R. L. and Ambros, V. (1993) 'The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14', *Cell*, 75, pp. 843–854.

Li, B., Dong, C., Li, P., Ren, Z., Wang, H., Yu, F., et al. (2016) 'Identification of candidate genes associated with porcine meat color traits by genome-wide transcriptome analysis', *Scientific Reports*, 6, p. 35224.

Li, W., Yang, Y., Liu, Y., Liu, S., Li, X., Wang, Y., et al. (2017) 'Integrated analysis of mRNA and miRNA expression profiles in livers of Yimeng black pigs with extreme phenotypes for backfat thickness', *Oncotarget*, 8(70), pp. 114787–114800.

Liang, Y., Wang, Y., Ma, L., Zhong, Z., Yang, X., Tao, X., et al. (2019) 'Comparison of microRNAs in adipose and muscle tissue from seven indigenous Chinese breeds and Yorkshire pigs', *Animal Genetics*, 50(5), pp. 439–448.

Liaubet, L., Lobjois, V., Faraut, T., Tircazes, A., Benne, F., Iannuccelli, N., et al. (2011) 'Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism', *BMC Genomics*, 12, p. 548.

Listrat, A., Lebret, B., Louveau, I., Astruc, T., Bonnet, M., Lefaucheur, L., et al. (2016) 'How muscle structure and composition influence meat and flesh quality', *The Scientific World Journal*, 2016, p. 3182746.

Liu, J., Damon, M., Guitton, N., Guisle, I., Ecolan, P., Vincent, A., et al. (2009) 'Differentially-expressed genes in pig longissimus muscles with contrasting levels of fat, as identified by combined transcriptomic, reverse transcription PCR, and proteomic analyses', *Journal of Agricultural and Food Chemistry*, 57(9), pp. 3808–3817.

Liu, X., Liu, H., Wang, M., Li, R., Zeng, J., Mo, D., et al. (2019) 'Disruption of the ZBED6 binding site in intron 3 of IGF2 by CRISPR/Cas9 leads to enhanced muscle development in Liang Guang Small Spotted pigs', *Transgenic Research*, 28(1), pp. 141–150.

López-Buesa, P., Burgos, C., Galve, A. and Varona, L. (2013) 'Joint analysis of additive , dominant and first-order epistatic effects of four genes (IGF2 , MC4R , PRKAG3

and LEPR) with known effects on fat content and fat distribution in pigs', *Animal Genetics*, 45, pp. 133–137.

López-Maury, L., Marguerat, S. and Bähler, J. (2008) 'Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation', *Nature Reviews Genetics*, 9(8), pp. 583–593.

Ma, N., Wang, X., Qiao, Y., Li, F., Hui, Y., Zou, C., et al. (2011) 'Coexpression of an intronic microRNA and its host gene reveals a potential role for miR-483-5p as an IGF2 partner', *Molecular and Cellular Endocrinology*, 333(1), pp. 96–101.

MacKay, T. F. C., Stone, E. A. and Ayroles, J. F. (2009) 'The genetics of quantitative traits: Challenges and prospects', *Nature Reviews Genetics*, 10(8), pp. 565–577.

Manunza, A., Casellas, J., Quintanilla, R., González-Prendes, R., Pena, R. N., Tibau, J., et al. (2014) 'A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants', *BMC Genomics*, 15(1), pp. 1–12.

Marquart, T. J., Allen, R. M., Ory, D. S. and Baldan, A. (2010) 'miR-33 links SREBP-2 induction to repression of sterol transporters', *Proceedings of the National Academy of Sciences*, 107(27), pp. 12228–12232.

Martínez-Montes, Á. M., Fernández, A., Muñoz, M., Noguera, J. L., Folch, J. M. and Fernández, A. I. (2018) 'Using genome wide association studies to identify common QTL regions in three different genetic backgrounds based on Iberian pig breed', *PLoS ONE*, 13(3), p. e0190184.

Martínez-Montes, A. M., Muiños-Bühl, A., Fernández, A., Folch, J. M., Ibáñez-Escriche, N. and Fernández, A. I. (2017) 'Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics', *Mammalian Genome*, 28(3–4), pp. 130–142.

Mata, J., Marguerat, S. and Bähler, J. (2005) 'Post-transcriptional control of gene expression: A genome-wide perspective', *Trends in Biochemical Sciences*, 30(9), pp. 506–514.

Mercadé, A., Estellé, J., Noguera, J. L., Folch, J. M., Varona, L., Silió, L., et al. (2005) 'On growth, fatness, and form: A further look at porcine Chromosome 4 in an Iberian x Landrace cross', *Mammalian Genome*, 16(5), pp. 374–382.

Mercadé, A., Pérez-Enciso, M., Varona, L., Alves, E., Noguera, J. L., Sánchez, A., et al. (2006) 'Adipocyte fatty-acid binding protein is closely associated to the porcine FAT1 locus on chromosome 4', *Journal of Animal Science*, 84(11), pp. 2907–2913.

Mercadé, A., Sánchez, A. and Folch, J. M. (2005) 'Exclusion of the acyl CoA:diacylglycerol acyltransferase 1 gene (DGAT1) as a candidate for a fatty acid composition QTL on porcine chromosome 4', *Journal of Animal Breeding and Genetics*, 122(3), pp. 161–164.

Mercadé, A., Sánchez, A. and Folch, J. M. (2007) 'Characterization and physical mapping of the porcine CDS1 and CDS2 genes', *Animal Biotechnology*, 18(1), pp. 23–35.

Meyer, C., Dostou, J. M., Welle, S. L. and Gerich, J. E. (2002) 'Role of human liver, kidney, and skeletal muscle in postprandial glucose homeostasis', *American Journal of Physiology-Endocrinology and Metabolism*, 282(2), pp. E419–E427.

Miar, Y., Plastow, G. and Wang, Z. (2015) 'Genomic Selection, a New Era for Pork Quality Improvement', *Springer Science Reviews*, 3(1), pp. 27–37.

Moll, P., Ante, M., Seitz, A. and Reda, T. (2014) 'QuantSeq 3 ' mRNA sequencing for RNA quantification', *Nature Methods*, 11, pp. 5–7.

Moon, Y., Shah, N. A., Mohapatra, S., Warrington, J. A. and Horton, J. D. (2001) 'Identification of a Mammalian Long Chain Fatty Acyl Elongase Regulated by Sterol Regulatory Element-binding Proteins', The *Journal of Biological Chemistry*, 276(48), pp. 45358–45366.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., et al. (2004) 'Genetic analysis of genome-wide variation in human gene expression', *Nature*, 430(7001), pp. 743–747.

Muñoz, G., Alves, E., Fernández, A., Óvilo, C., Barragán, C., Estellé, J., et al. (2007) 'QTL detection on porcine chromosome 12 for fatty-acid composition and association analyses of the fatty acid synthase, gastric inhibitory polypeptide and acetyl-coenzyme A carboxylase alpha genes', *Animal Genetics*, 38(6), pp. 639–646.

Muñoz, G., Ovilo, C., Silló, L., Tomás, A., Noguera, J. L. and Rodríguez, M. C. (2009) 'Single- And joint-population analyses of two experimental pig crosses to confirm quantitative trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits', *Journal of Animal Science*, 87(2), pp. 459–468.

Muñoz, M., Bozzi, R., García, F., Núñez, Y., Geraci, C., Crovetti, A., et al. (2018a) 'Diversity across major and candidate genes in European local pig breeds', *PLoS ONE*, 13(11), p. e0207475.

Muñoz, M., García-Casco, J. M., Caraballo, C., Fernández-Barroso, M. Á., Sánchez-

Esquiliche, F., Gómez, F., et al. (2018b) 'Identification of Candidate Genes and Regulatory Factors Underlying Intramuscular Fat Content Through Longissimus Dorsi Transcriptome Analyses in Heavy Iberian Pigs', *Frontiers in Genetics*, 9(12), pp. 1–16.

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

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

Myers, A. J., Gibbs, J. R., Webster, J. A., Rohrer, K., Zhao, A., Marlowe, L., et al. (2007) 'A survey of genetic human cortical gene expression', *Nature Genetics*, 39(12), pp. 1494–1499.

Najafi-Shoushtari, S., Kristo, F., Li, Y., Shioda, T., Cohen, D., Gerszten, R., et al. (2010) 'MicroRNA-33 and the SREBP Host Genes Cooperate to Control Cholesterol Homeostasis', *Science*, 328(405), pp. 57–65.

Nakamura, M. T. and Nara, T. Y. (2004) 'Structure, Function, and Dietary Regulation of Δ6, Δ5, and Δ9 Desaturases', *Annual Review of Nutrition*, 24(1), pp. 345–376.

Neve, B., Froguel, P., Corset, L., Vaillant, E., Vatin, V. and Boutin, P. (2002) 'Rapid SNP allele frequency determination in genomic DNA pools by PyrosequencingTM', *BioTechniques*, 32(5), pp. 1138–1142.

Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., et al. (1999) 'An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs.', *Nature Genetics*, 21(2), pp. 155–156.

Nguyen, P., Leray, V., Diez, M., Serisier, S., Le Bloc'H, J., Siliart, B., et al. (2008) 'Liver lipid metabolism', *Journal of Animal Physiology and Animal Nutrition*, 92(3), pp. 272–283.

Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlén, M., et al. (2012) 'A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: A case study in Saccharomyces cerevisiae', *Nucleic Acids Research*, 40(20), pp. 10084–10097.

O'Brien, J., Hayder, H., Zayed, Y. and Peng, C. (2018) 'Overview of microRNA

biogenesis, mechanisms of actions, and circulation', *Frontiers in Endocrinology*, 9, pp. 1–12.

O'Connell, M. J., Walworth, N. C. and Carr, A. M. (2000) 'The G2-phase DNA-damage checkpoint', Trends in Cell Biology, 10(7), pp. 296–303.

O'Hea, E. K. and Leveille, G. A. (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(3), pp. 338–344.

Ohashi, H., Hasegawa, M., Wakimoto, K. and Miyamoto-Sato, E. (2015) 'Next-generation technologies for multiomics approaches including interactome sequencing', *BioMed Research International*, 2015, p. 104209.

Óvilo, C., Benítez, R., Fernández, A., Núñez, Y., Ayuso, M., Fernández, A. I., et al. (2014) 'Longissimus dorsi transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics', *BMC Genomics*, 15, p. 413.

Óvilo, C., Clop, A., Noguera, J. L., Oliver, M. A., Barragán, C., Rodríguez, C., et al. (2002) 'Quantitative trait locus mapping for meat quality traits in an Iberian x Landrace F2 pig population', *Journal of Animal Science*, 80(11), pp. 2801–2808.

Óvilo, C., Fernández, A., Noguera, J. L., Barragán, C., Letón, R., Rodríguez, C., et al. (2005) 'Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross', *Genetical Research*, 85(1), pp. 57–67.

Óvilo, C., Pérez-Enciso, M., Barragán, C., Clop, A., Rodríguez, C., Angels Oliver, M., et al. (2000) 'A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6', *Mammalian Genome*, 11(4), pp. 344–346.

Pena, R. N., Noguera, J. L., Casellas, J., Díaz, I., Fernández, A. I., Folch, J. M., et al. (2013) 'Transcriptional analysis of intramuscular fatty acid composition in the longissimus thoracis muscle of Iberian × Landrace back-crossed pigs', *Animal Genetics*, 44(6), pp. 648–660.

Pérez-Enciso, M., Clop, A., Folch, J. M., Sánchez, A., Oliver, M. A., Óvilo, C., et al. (2002) 'Exploring alternative models for sex-linked quantitative trait loci in outbred populations: Application to an iberian × landrace pig intercross', *Genetics*, 161(4), pp. 1625–1632.

Pérez-Enciso, M., Clop, A., Noguera, J. L., Ovilo, C., Coll, A., Folch, J. M., et al. (2000) 'A QTL on pig chromosome 4 affects fatty acid metabolism : evidence from an Iberian by Landrace intercross', *Journal of Animal Science*, 78(10), pp. 2525–2531.

Pérez-Enciso, M., Mercadé, A., Bidanel, J. P., Geldermann, H., Cepica, S., Bartenschlager, H., et al. (2005) 'Large-scale, multibreed, multitrait analyses of quantitative trait loci experiments: The case of porcine X chromosome', *Journal of Animal Science*, 83(10), pp. 2289–2296.

Pérez-Enciso, M., Rincón, J. C. and Legarra, A. (2015) 'Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised', *Genetics Selection Evolution*, 47(1), pp. 1–14.

Pérez-Montarelo, D., Hudson, N. J., Fernández, A. I., Ramayo-Caldas, Y., Dalrymple, B. P. and Reverter, A. (2012) 'Porcine Tissue-Specific Regulatory Networks Derived from Meta-Analysis of the Transcriptome', *PLoS ONE*, 7(9), p. e46159.

Pfaffl, M. W. (2012) 'Quantification Strategies in Real-time Polymerase Chain Reaction', *Polymerase Chain Reaction: Theory and Technology*. Chapter 3, pp. 105-114.

Pierce, B. L., Tong, L., Chen, L. S., Rahaman, R., Argos, M., Jasmine, F., et al. (2014) 'Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians', *PLoS Genetics*, 10(12), p. e1004818.

Piórkowska, K., Żukowski, K., Ropka-Molik, K., Tyra, M. and Gurgul, A. (2018) 'A comprehensive transcriptome analysis of skeletal muscles in two Polish pig breeds differing in fat and meat quality traits', *Genetics and Molecular Biology*, 41(1), pp. 125–136.

Ponsuksili, S., Du, Y., Murani, E., Schwerin, M. and Wimmers, K. (2012) 'Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle', *Genetics*, 192(3), pp. 1109–1122.

Ponsuksili, S., Jonas, E., Murani, E., Phatsara, C., Srikanchai, T., Walz, C., et al. (2008) 'Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle', *BMC Genomics*, 9, p. 367.

Ponsuksili, S., Murani, E., Brand, B., Schwerin, M. and Wimmers, K. (2011) 'Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model', *Journal of Lipid Research*, 52(4), pp. 668–678.

Ponsuksili, S., Murani, E., Schwerin, M., Schellander, K. and Wimmers, K. (2010) 'Identification of expression QTL (eQTL) of genes expressed in porcine M. longissimus dorsi and associated with meat quality traits', *BMC Genomics*, 11, p. 572.

Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M. and Wimmers, K. (2014) 'Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis', *International Journal of Biological Sciences*, 10(3), pp. 327–337.

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D., Hudson, N. J., et al. (2014) 'Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition', *PLoS ONE*, 9(6), p. e99720.

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A. I., Folch, J. M. and Ballester, M. (2016) 'Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat.', *Scientific Reports*, 6(2), p. 31803.

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., et al. (2012a) 'Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition', *BMC Genomics*, 13, p. 547.

Ramayo-Caldas, Y., Mercade, A., Castello, A., Yang, B., Rodriguez, C., Alves, E., et al. (2012b) 'Genome-wide association study for intramuscular fatty acid composition in an Iberian x Landrace cross', *Journal of Animal Science*, 90(9), pp. 2883–2893.

Ramirez, C. M., Goedeke, L., Rotllan, N., Yoon, J.-H., Cirera-Salinas, D., Mattison, J. A., et al. (2013) 'MicroRNA 33 Regulates Glucose Metabolism', *Molecular and Cellular Biology*, 33(15), pp. 2891–2902.

Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., et al. (2009) 'Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology', *PLoS ONE*, 4(8), p. e6524.

Rayner, K. J., Suárez, Y., Dávalos, A., Parathath, S., Michael, L., Tamehiro, N., et al. (2010) 'MiR-33 Contributes to the Regulation of Cholesterol Homeostasis', *Science*, 328(5985), pp. 1570–1573.

Reddy, A. M., Zheng, Y., Jagadeeswaran, G., Macmil, S. L., Graham, W. B., Roe, B. A., et al. (2009) 'Cloning, characterization and expression analysis of porcine microRNAs', *BMC Genomics*, 10, p. 65.

Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., Criado-Mesas, L., Castelló, A., Fernández, A. I., et al. (2018) 'Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions', *Scientific Reports*, 8(1), pp. 1–13.

Revilla, M., Ramayo-Caldas, Y., Castelló, A., Corominas, J., Puig-Oliveras, A., Ibáñez-

Escriche, N., et al. (2014) 'New insight into the SSC8 genetic determination of fatty acid composition in pigs', *Genetics Selection Evolution*, 46(1), pp. 1–10.

Ricroch, A. (2019) 'Global developments of genome editing in agriculture', *Transgenic Research*, 28, pp. 45–52.

Sanger, F., Nicklen, S. and Coulson, R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proceedings of the National Academy of Sciences at United States of America*, 74(12), pp. 5463-5467.

Sawera, M., Gorodkin, J., Cirera, S. and Fredholm, M. (2005) 'Mapping and expression studies of the mir17-92 cluster on pig Chromosome 11', *Mammalian Genome*, 16(8), pp. 594–598.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., et al. (2003) 'Genetics of gene expression surveyed in maize, mouse and man', *Nature*, 422(6929), pp. 297–302.

Schook, L. B., Beever, J. E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., et al. (2005) 'Swine Genome Sequencing Consortium (SGSC): A strategic roadmap for sequencing the pig genome', *Comparative and Functional Genomics*, 6(4), pp. 251–255.

Serra, X., Gil, F., Pérez-Enciso, M., Oliver, M. A., Vázquez, J. M., Gispert, M., et al. (1998) 'A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs', *Livestock Production Science*, 56(3), pp. 215–223.

Sharmaa, A., Lee, J. S., Dang, C. G., Sudrajad, P., Kim, H. C., Yeon, S. H., et al. (2015) 'Stories and challenges of genome wide association studies in livestock - a review', *Asian-Australasian Journal of Animal Sciences*, 28(10), pp. 1371–1379.

Shimano, H. (2001) 'Sterol Regulatory Element-Binding Proteins (SREBPs) as Regulators of Lipid Metabolism', *Progress in Lipid Research*, (40), pp. 439–452.

Sodhi, S. S., Song, K. D., Ghosh, M., Sharma, N., Lee, S. J., Kim, J. H., et al. (2014) 'Comparative transcriptomic analysis by RNA-seq to discern differential expression of genes in liver and muscle tissues of adult Berkshire and Jeju Native Pig', *Gene*, 546(2), pp. 233–242.

Song, Z., Cooper, D. K. C., Cai, Z. and Mou, L. (2018) 'Expression and regulation profile of mature microRNA in the pig: Relevance to xenotransplantation', *BioMed Research International*, 2018, p. 2983908.

Southern, E. M. (2001) 'DNA microarrays. History and overview.', *Methods Molecular Biology*, 170(4), pp. 1–15.

Spurgeon, S. L., Jones, R. C. and Ramakrishnan, R. (2008) 'High throughput gene expression measurement with real time PCR in a microfluidic dynamic array', *PLoS ONE*, 3(2), p. e1662.

Steibel, J. P., Bates, R. O., Rosa, G. J. M., Tempelman, R. J., Rilington, V. D., Ragavendran, A., et al. (2011) 'Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs', *PLoS ONE*, 6(2), p. 16766.

Steibel, J. P., Wysocki, M., Lunney, J. K., Ramos, A. M., Hu, Z. L., Rothschild, M. F., et al. (2009) 'Assessment of the swine protein-annotated oligonucleotide microarray', *Animal Genetics*, 40(6), pp. 883–893.

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J. and Friess, H. (2009) 'Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues', *Nature Protocols*, 4(1), pp. 37–43.

Sun, W. X., Wang, H. H., Jiang, B. C., Zhao, Y. Y., Xie, Z. R., Xiong, K., et al. (2013) 'Global comparison of gene expression between subcutaneous and intramuscular adipose tissue of mature Erhualian pig', *Genetics and Molecular Research*, 12(4), pp. 5085–5101.

Sun, W., Yu, T. and Li, K. (2007) 'Detection of eQTL modules mediated by activity levels of transcription factors', *Bioinformatics*, 23(17), pp. 2290–2297.

Suravajhala, P., Kogelman, L. J. A. and Kadarmideen, H. N. (2016) 'Multi-omic data integration and analysis using systems genomics approaches: Methods and applications In animal production, health and welfare', *Genetics Selection Evolution*, 48(1), pp. 1–14.

Suzuki, K., Irie, M., Kadowaki, H., Shibata, T., Kumagai, M. and Nishida, A. (2005) 'Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content', *Journal of Animal Science*, 83(9), pp. 2058–2065.

Taniguchi, M., Arakawa, A., Motoyama, M., Nakajima, I., Nii, M. and Mikawa, S. (2015) 'Genomic structural analysis of porcine fatty acid desaturase cluster on chromosome 2', *Animal Science Journal*, 86(4), pp. 369–377.

Taniguchi, M., Nakajima, I., Chikuni, K., Kojima, M., Awata, T. and Mikawa, S. (2014) 'MicroRNA-33b downregulates the differentiation and development of porcine preadipocytes', *Molecular Biology Reports*, 41(2), pp. 1081–1090.

Tsai, A. C., Romsos, D. R. and Leveille, G. A. (1975) 'Effect of Dietary Cholesterol on Hepatic Lipogenesis and Plasma Insulin and Free Fatty Acid Levels in Rats', *The*

Journal of Nutrition, 105(7), pp. 939–945.

Van Laere, A. S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., et al. (2003) 'A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.', *Nature*, 425(10), pp. 832–836.

Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) 'Serial analysis of gene expression (SAGE).', *Methods in molecular biology*, 270(10), pp. 41–66.

Velez-Irizarry, D., Casiro, S., Daza, K. R., Bates, R. O., Raney, N. E., Steibel, J. P., et al. (2019) 'Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs 06 Biological Sciences 0604 Genetics', *BMC Genomics*, 20(1), pp. 1–19.

Verdugo, R. A., Farber, C. R., Warden, C. H. and Medrano, J. F. (2010) 'Serious limitations of the QTL/Microarray approach for QTL gene discovery', *BMC Biology*, 8, p. 96.

Wang, J., Vasaikar, S., Shi, Z., Greer, M. and Zhang, B. (2017) 'WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit', *Nucleic Acids Research*, 45(W1), pp. W130–W137.

Wang, W. Y. S., Barratt, B. J., Clayton, D. G. and Todd, J. A. (2005) 'Genome-wide association studies: Theoretical and practical concerns', *Nature Reviews Genetics*, 6(2), pp. 109–118.

Wang, X., Gu, Z. and Jiang, H. (2013) 'MicroRNAs in farm animals', *Animal*, 7(10), pp. 1567–1575.

Wasson, J., Skolnick, G., Love-Gregory, L. and Permutt, M. A. (2002) 'Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by PyrosequencingTM technology', *BioTechniques*, 32(5), pp. 1144–1152.

Webb, E. C. and O'Neill, H. A. (2008) 'The animal fat paradox and meat quality', *Meat Science*, 80(1), pp. 28–36.

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013) 'Systematic identification of trans eQTLs as putative drivers of known disease associations', *Nature Genetics*, 45(10), pp. 1238–1243.

Wightman, B., Ha, I. and Ruvkun, G. (1993) 'Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans', *Cell*, 75(5), pp. 855–862.

Wimmers, K., Murani, E. and Ponsuksili, S. (2010) 'Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat

performance in pigs', *Briefings in Functional Genomics and Proteomics*, 9(3), pp. 251–258.

Won, S., Jung, J., Park, E. and Kim, H. (2018) 'Identification of genes related to intramuscular fat content of pigs using genome-wide association study', *Asian-Australasian Journal of Animal Sciences*, 31(2), pp. 157–162.

Wood, J. D. and Enser, M. (1997) 'Factors influencing fatty acids in meat and the role of antioxidants in improving meat quality', *British Journal of Nutrition*, 78(1), pp. S49–S60.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Richardson, R. I. and Sheard, P. R. (1999) 'Manipulating meat quality and composition', *Proceedings of the Nutrition Society*, 58(2), pp. 363–370.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., et al. (2008) 'Fat deposition, fatty acid composition and meat quality: A review', *Meat Science*, 78(4), pp. 343–358.

Wood, J. D., Richardson, R. I., Nute, G. R., Fisher, A. V., Campo, M. M., Kasapidou, E., et al. (2004) 'Effects of fatty acids on meat quality: A review', *Meat Science*, 66(1), pp. 21–32.

Wood, J. and Whittemore, C. (2007) 'Pig Meat and Carcass Quality', *Whittemore's Science and Practice of Pig Production*. 3rd edn. Oxford, UK: Blackwell Publishing Ltd.

Wu, G. and Bazer, F. W. (2019) 'Application of new biotechnologies for improvements in swine nutrition and pork production', *Journal of Animal Science and Biotechnology*, 10(1), pp. 1–16.

Xing, K., Wang, K., Ao, H., Chen, S., Tan, Z., Wang, Y., et al. (2019a) 'Comparative adipose transcriptome analysis digs out genes related to fat deposition in two pig breeds', *Scientific Reports*, 9(1), pp. 1–11.

Xing, K., Zhao, X., Ao, H., Chen, S., Yang, T., Tan, Z., et al. (2019b) 'Transcriptome analysis of miRNA and mRNA in the livers of pigs with highly diverged backfat thickness', *Scientific Reports*, 9(1), pp. 1–12.

Xing, K., Zhu, F., Zhai, L., Liu, H., Wang, Z., Hou, Z., et al. (2014) 'The liver transcriptome of two full-sibling Songliao black pigs with extreme differences in backfat thickness', *Journal of Animal Science* and Biotechnology, 5(1), pp. 1–9.

Yang, H. and Wu, Z. (2018) 'Genome Editing of Pigs for Agriculture and Biomedicine', *Frontiers in Genetics*, 9, p. 360.

Yao, C., Joehanes, R., Johnson, A. D., Huan, T., Liu, C., Freedman, J. E., et al. (2017) 'Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits', *American Journal of Human Genetics*, 100(4), pp. 571–580.

Yin, L., Ma, H., Ge, X., Edwards, P. A. and Zhang, Y. (2011) 'Hepatic hepatocyte nuclear factor 4α is essential for maintaining triglyceride and cholesterol homeostasis', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(2), pp. 328–336.

Yu, K., Shu, G., Yuan, F., Zhu, X., Gao, P., Wang, S., et al. (2013) 'Fatty acid and transcriptome profiling of longissimus dorsi muscles between pig breeds differing in meat quality', *International Journal of Biological Sciences*, 9(1), pp. 108–118.

Zhang, C. Y., Wang, Z., Bruce, H. L., Janz, J., Goddard, E., Moore, S., et al. (2014) 'Associations between single nucleotide polymorphisms in 33 candidate genes and meat quality traits in commercial pigs', *Animal Genetics*, 45(4), pp. 508–516.

Zhang, H. Z., Chen, D. W., He, J., Zheng, P., Yu, J., Mao, X. B., et al. (2019) 'Long-term dietary resveratrol supplementation decreased serum lipids levels, improved intramuscular fat content, and changed the expression of several lipid metabolism-related miRNAs and genes in growing-finishing pigs', *Journal of Animal Science*, 97(4), pp. 1745–1756.

Zhao, Chen, Tan, Wang, Zhang, Yang, et al. (2019) 'Transcriptome Analysis of Landrace Pig Subcutaneous Preadipocytes during Adipogenic Differentiation', *Genes*, 10(7), p. 552.

Zhao, X., Mo, D., Li, A., Gong, W., Xiao, S., Zhang, Y., et al. (2011) 'Comparative analyses by sequencing of transcriptomes during skeletal muscle development between pig breeds differing in muscle growth rate and fatness', *PLoS ONE*, 6(5), p. e19774.

Zhao, Z. D., Zan, L. Sen, Li, A. N., Cheng, G., Li, S. J., Zhang, Y. R., et al. (2016) 'Characterization of the promoter region of the bovine long-chain acyl-CoA synthetase 1 gene: Roles of E2F1, Sp1, KLF15, and E2F4', *Scientific Reports*, 6, pp. 1–9.

Annexes

Chapter 7

7.1. Supplementary material Paper I: 'Identification of eQTLs associated with lipid metabolism in *Longissimus dorsi* muscle of pigs with different genetic backgrounds'

Table S1. List of significant associated SNPs within eQTLs intervals for the 45-muscle gene expression study in 3BCs.

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
1	3	ACSM5.P	25,422	0.133	1.39E-27	5.34E-23	ACSM5	-	-	
1	3	rs81278505	25,651	0.117	1.69E-24	3.24E-20	ACSM5	-	-	
1	3	rs81227560	23,359	0.145	9.45E-22	1.21E-17	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
1	3	rs81312070	23,335	0.856	2.40E-21	2.30E-17	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
1	3	rs81475068	23,968	0.936	7.29E-15	5.60E-11	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
1	3	rs81347321	25,606	0.271	1.93E-13	1.24E-09	ACSM5	-	-	-
1	3	rs81340946	23,582	0.157	2.51E-13	1.38E-09	ACSM5	intron variant	ENSSSCG00000007838	OTOA
1	3	rs81238437	27,259	0.081	2.46E-12	1.18E-08	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
1	3	rs81313849	23,887	0.788	2.80E-12	1.19E-08	ACSM5	upstream gene variant	ENSSSCG00000020312	RF00026
1	3	rs81239835	26,750	0.915	1.56E-11	5.99E-08	ACSM5	intron variant	ENSSSCG00000032531	SMG1
1	3	rs81309174	22,744	0.842	3.65E-11	1.27E-07	ACSM5	intron variant	ENSSSCG00000010799	COG7
1	3	rs81379272	25,852	0.614	9.64E-11	3.09E-07	ACSM5	-	-	-
1	3	rs81240993	25,654	0.191	4.14E-10	1.22E-06	ACSM5	-	-	-
1	3	rs81238947	23,519	0.243	9.89E-10	2.71E-06	ACSM5	-	-	-
1	3	rs81313219	22,753	0.174	3.17E-09	8.13E-06	ACSM5	intron variant	ENSSSCG00000010799	COG7
1	3	rs81324887	23,568	0.774	6.09E-09	1.46E-05	ACSM5	downstream gene variant	ENSSSCG00000007838	OTOA
1	3	rs81335959	25,260	0.248	1.39E-08	2.88E-05	ACSM5	intron variant	ENSSSCG00000007857	ACSM3
1	3	rs81315362	25,206	0.753	1.42E-08	2.88E-05	ACSM5	intron variant	ENSSSCG00000007855	REXO5
1	3	rs81288413	25,217	0.247	1.42E-08	2.88E-05	ACSM5	intron variant	ENSSSCG00000007855	REXO5
1	3	rs81322563	18,557	0.955	2.62E-08	4.95E-05	ACSM5	intron variant	ENSSSCG00000007805	ATP2A1
1	3	rs81326933	25,540	0.76	2.71E-08	4.95E-05	ACSM5	-	-	-
1	3	rs81379199	25,155	0.749	3.02E-08	5.27E-05	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
1	3	rs81379197	25,173	0.252	4.12E-08	6.88E-05	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
1	3	rs81288253	24,942	0.736	4.89E-08	7.83E-05	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
1	3	rs81326798	24,892	0.744	6.03E-08	9.26E-05	ACSM5	intron variant	ENSSSCG00000031876	TMEM159
1	3	rs81308074	23,896	0.262	1.91E-07	2.68E-04	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
1	3	rs323881880	23,900	0.261	1.92E-07	2.68E-04	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
1	3	rs324741666	26,460	0.849	2.02E-07	2.68E-04	ACSM5	intron variant	ENSSSCG00000022200	SYT17
1	3	rs81474976	26,514	0.849	2.02E-07	2.68E-04	ACSM5	intron variant	ENSSSCG00000022200	SYT17
1	3	rs81475002	26,275	0.867	2.93E-07	3.76E-04	ACSM5	intron variant	ENSSSCG00000007868	TMC5
1	3	rs81330380	23,056	0.778	5.45E-07	6.76E-04	ACSM5	intron variant	ENSSSCG00000030424	USP31
1	3	rs80876065	23,093	0.223	6.26E-07	7.51E-04	ACSM5	downstream gene variant	ENSSSCG00000023632	RF00414
1	3	rs81379308	27,015	0.464	6.65E-07	7.74E-04	ACSM5	-	-	-
1	3	rs81335819	23,818	0.53	7.31E-07	8.17E-04	ACSM5	intron variant	ENSSSCG00000032340	-
1	3	rs81309807	22,735	0.196	7.44E-07	8.17E-04	ACSM5	downstream gene variant	ENSSSCG00000010798	GGA2
1	3	rs81321464	23,166	0.35	1.15E-06	1.23E-03	ACSM5	-	-	-
1	3	rs81293818	22,810	0.936	1.19E-06	1.23E-03	ACSM5	intron variant	ENSSSCG00000010799	COG7
1	3	rs81318451	20,859	0.194	1.78E-06	1.80E-03	ACSM5	-	-	-
1	3	rs81324453	47,279	0.071	2.23E-06	2.19E-03	ACSM5	intron variant	ENSSSCG00000038296	SH3RF3
1	3	rs81336877	24,920	0.665	3.07E-06	2.95E-03	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
1	3	rs81278892	24,815	0.233	3.69E-06	3.45E-03	ACSM5	5 prime UTR variant	ENSSSCG00000007849	CRYM

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
1	3	rs81329230	26,346	0.571	4.52E-06	4.14E-03	ACSM5	intron variant	ENSSSCG00000007866	TMC7
1	3	rs81311562	22,718	0.747	1.47E-05	1.31E-02	ACSM5	intron variant	ENSSSCG00000010798	GGA2
1	3	rs81306471	22,905	0.216	1.52E-05	1.33E-02	ACSM5	upstream gene variant	ENSSSCG00000028923	SCNN1B
1	3	rs81289409	36,102	0.076	1.90E-05	1.63E-02	ACSM5	-	-	-
1	3	rs81379171	25,072	0.819	2.06E-05	1.72E-02	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
1	3	rs81314488	22,771	0.181	2.26E-05	1.85E-02	ACSM5	intron variant	ENSSSCG00000010799	COG7
1	3	rs81323675	24,582	0.109	2.64E-05	2.08E-02	ACSM5	intron variant	ENSSSCG00000034298	-
1	3	rs81315383	24,980	0.749	3.25E-05	2.46E-02	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
1	3	rs81244431	25,758	0.168	3.26E-05	2.46E-02	ACSM5	downstream gene variant	ENSSSCG00000007862	GPR139
1	3	rs81234875	24,529	0.936	3.39E-05	2.50E-02	ACSM5	-	-	-
1	3	rs81477531	22,623	0.404	4.02E-05	2.79E-02	ACSM5	intron variant	ENSSSCG00000010795	-
1	3	rs81284839	22,634	0.596	4.02E-05	2.79E-02	ACSM5	downstream gene variant	ENSSSCG00000010795	-
1	3	rs81249771	25,004	0.565	4.06E-05	2.79E-02	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
1	3	rs81379431	27,322	0.15	4.59E-05	3.09E-02	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
1	3	rs81475137	23,446	0.261	4.92E-05	3.26E-02	ACSM5	-	-	-
1	3	rs81344302	23,621	0.476	6.94E-05	4.52E-02	ACSM5	intron variant	ENSSSCG00000007838	OTOA
1	3	rs81370592	53,699	0.773	7.75E-05	4.96E-02	ACSM5	-	-	-
2	6	rs81216702	17,315	0.515	1.38E-05	4.43E-02	ACSS2	synonymous variant	ENSSSCG00000002755	NFAT5
2	6	rs81246307	17,502	0.485	1.38E-05	4.43E-02	ACSS2	-	-	-
3	7	rs80870743	112,227	0.192	3.91E-06	4.06E-02	ACSS2	-	-	-
3	7	rs80913379	111,283	0.634	5.21E-06	4.06E-02	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
3	7	rs80871598	111,558	0.192	8.44E-06	4.06E-02	ACSS2	-	-	-
3	7	rs81396214	111,780	0.808	8.44E-06	4.06E-02	ACSS2	intron variant	ENSSSCG00000002431	TDP1
3	7	rs81396246	111,869	0.808	8.44E-06	4.06E-02	ACSS2	intron variant	ENSSSCG00000002432	KCNK13
3	7	rs80938538	111,492	0.375	1.02E-05	4.37E-02	ACSS2	-	-	-
3	7	rs81001496	112,194	0.193	1.18E-05	4.43E-02	ACSS2	-	-	-
3	7	rs80839580	112,044	0.193	1.53E-05	4.52E-02	ACSS2	3 prime UTR variant	ENSSSCG00000002433	PSMC1
4	13	rs80786918	156,644	0.68	7.48E-06	4.06E-02	ACSS2	intron variant	ENSSSCG00000011947	ZPLD1
4	13	rs80868545	156,576	0.335	8.13E-06	4.06E-02	ACSS2	-	-	-
5	1	rs80868279	181,702	0.637	8.60E-08	1.51E-03	ATF3	synonymous variant	ENSSSCG00000005030	NID2
5	1	rs80863162	181,624	0.359	5.24E-07	3.97E-03	ATF3	downstream gene variant	ENSSSCG00000035178	-
5	1	rs80962176	181,648	0.359	5.24E-07	3.97E-03	ATF3	intron variant	ENSSSCG00000035178	-
6	13	rs81478407	177,313	0.289	1.13E-07	1.51E-03	ATF3	-	-	-
6	13	rs81344735	177,546	0.723	6.20E-07	3.97E-03	ATF3	intron variant	ENSSSCG00000012002	ROBO2
7	16	rs81457359	2,764	0.38	3.58E-07	1.37E-02	DGAT2	-	-	-
7	16	rs81457374	2,779	0.19	3.63E-06	4.65E-02	DGAT2	-	-	-
8	1	rs80813421	261	0.124	1.76E-06	1.35E-02	FOS	-	-	-
8	1	rs80803041	493	0.126	2.36E-06	1.38E-02	FOS	-	-	-
9	11	rs80845358	10,367	0.934	2.30E-10	8.83E-06	FOS	-	-	-
9	11	rs81430187	19,677	0.636	1.02E-06	1.03E-02	FOS	-	-	-
9	11	rs80796231	8,855	0.15	2.52E-06	1.38E-02	FOS	intron variant	ENSSSCG00000029039	BRCA2
10	2	IGF2	1,483	0.424	3.24E-44	1.24E-39	IGF2	-	-	-
10	2	rs81306755	145	0.37	5.47E-39	1.05E-34	IGF2	intron variant	ENSSSCG00000014565	-
10	2	rs81317307	310	0.415	4.91E-37	6.29E-33	IGF2	intron variant	ENSSSCG00000027045	LRRC56
10	2	rs81357266	3,985	0.362	9.98E-32	9.59E-28	IGF2	-	-	-
10	2	rs81341763	677	0.453	3.63E-30	2.79E-26	IGF2	-	-	-
10	2	rs81328276	236	0.451	1.85E-28	1.19E-24	IGF2	intron variant	ENSSSCG00000024569	ANO9
10	2	rs81339115	422	0.476	3.50E-24	1.92E-20	IGF2	intron variant	ENSSSCG00000012850	DEAF1

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
10	2	rs81364067	4,412	0.565	8.34E-21	4.01E-17	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
10	2	rs81364734	4,444	0.563	5.74E-20	2.45E-16	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
10	2	rs81360111	8,647	0.561	1.09E-16	4.18E-13	IGF2	intron variant	ENSSSCG00000028537	-
10	2	rs81322199	3,689	0.161	2.18E-16	7.62E-13	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81291529	2,636	0.444	3.45E-16	1.10E-12	IGF2	intron variant	ENSSSCG00000033043	SHANK2
10	2	rs81336288	3,062	0.707	8.24E-16	2.44E-12	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81363333	4,378	0.439	1.71E-15	4.71E-12	IGF2	-	-	-
10	2	rs81358530	6,962	0.597	5.23E-15	1.34E-11	IGF2	intron variant	ENSSSCG00000012999	CAPN1
10	2	rs81356987	3,859	0.628	5.30E-14	1.27E-10	IGF2	-	-	-
10	2	rs81356358	5,289	0.427	5.55E-13	1.25E-09	IGF2	intron variant	ENSSSCG00000029637	KDM2A
10	2	rs81474931	11,764	0.269	5.00E-12	1.07E-08	IGF2	non coding transcript exon variant	ENSSSCG00000037095	-
10	2	rs81361514	4,341	0.416	7.66E-12	1.55E-08	IGF2	intron variant	ENSSSCG00000032760	-
10	2	rs81341267	3,094	0.655	1.08E-11	2.08E-08	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81238148	11,975	0.573	1.41E-10	2.57E-07	IGF2	upstream gene variant	ENSSSCG00000031679	-
10	2	rs81359966	8,687	0.693	1.56E-10	2.73E-07	IGF2	intron variant	ENSSSCG00000028537	-
10	2	rs81368683	4,966	0.331	4.03E-10	6.73E-07	IGF2	upstream gene variant	ENSSSCG00000012896	NDUFV1
10	2	rs81252426	2,984	0.614	1.45E-09	2.33E-06	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81357172	6,263	0.131	2.30E-09	3.54E-06	IGF2	-	-	-
10	2	rs81237341	4,671	0.076	4.47E-09	6.61E-06	IGF2	-	-	-
10	2	rs81330112	3,058	0.627	4.92E-09	7.00E-06	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81343851	2,149	0.347	5.76E-09	7.90E-06	IGF2	intron variant	ENSSSCG00000012857	CARS
10	2	rs81257178	11,836	0.77	3.22E-08	4.27E-05	IGF2	intron variant	ENSSSCG00000013124	PATL1
10	2	rs81360839	10,031	0.369	4.62E-08	5.91E-05	IGF2	-	-	-
10	2	rs81362513	12,438	0.513	5.72E-08	7.09E-05	IGF2	intron variant	ENSSSCG00000032136	-
10	2	rs333411238	3,257	0.414	5.94E-08	7.14E-05	IGF2	intron variant	ENSSSCG00000031191	-
10	2	rs81214179	8,936	0.632	6.16E-08	7.17E-05	IGF2	upstream gene variant	ENSSSCG00000026293	STX5
10	2	rs81345516	12,209	0.125	8.02E-08	9.06E-05	IGF2	intron variant	ENSSSCG00000013145	DTX4
10	2	rs81285409	8,393	0.455	1.16E-07	1.27E-04	IGF2	upstream gene variant	ENSSSCG00000040120	-
10	2	rs81340329	13,620	0.261	1.20E-07	1.28E-04	IGF2	upstream gene variant	ENSSSCG00000029468	P2RX3
10	2	rs81360021	8,765	0.628	1.25E-07	1.30E-04	IGF2	-	-	-
10	2	rs81294446	5,339	0.19	1.30E-07	1.31E-04	IGF2	intron variant	ENSSSCG00000029637	KDM2A
10	2	rs81346312	5,372	0.21	1.51E-07	1.49E-04	IGF2	-	-	-
10	2	rs81361507	11,440	0.379	1.57E-07	1.51E-04	IGF2	-	-	-
10	2	rs81474907	8,795	0.776	2.24E-07	2.10E-04	IGF2	intron variant	ENSSSCG00000029516	SLC22A8
10	2	rs81285769	5,830	0.178	2.33E-07	2.14E-04	IGF2	upstream gene variant	ENSSSCG00000012933	CCDC87
10	2	rs81333729	2,380	0.286	3.15E-07	2.82E-04	IGF2	intron variant	ENSSSCG00000021181	DHCR7
10	2	rs81359337	7,852	0.72	3.45E-07	3.01E-04	IGF2	intron variant	ENSSSCG00000034755	-
10	2	rs81362233	10,687	0.363	3.55E-07	3.03E-04	IGF2	-	-	-
10	2	rs81314686	9,226	0.163	4.39E-07	3.66E-04	IGF2	intron variant	ENSSSCG00000036669	-
10	2	rs81271991	6,640	0.275	4.48E-07	3.66E-04	IGF2	upstream gene variant	ENSSSCG00000012983	PCNX3
10	2	rs81330355	8,819	0.677	5.14E-07	4.11E-04	IGF2	downstream gene variant	ENSSSCG00000023571	SLC22A6
10	2	rs341817021	3,978	0.599	6.95E-07	5.45E-04	IGF2	-	-	-
10	2	rs81360570	9,772	0.746	7.20E-07	5.53E-04	IGF2	intron variant	ENSSSCG00000013078	MYRF
10	2	rs332366314	13,156	0.501	7.97E-07	6.01E-04	IGF2	3 prime UTR variant	ENSSSCG00000013174	CTNND1
10	2	rs81361464	10,543	0.241	9.30E-07	6.87E-04	IGF2	intron variant	ENSSSCG00000029938	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
10	2	rs81322752	12,817	0.477	9.79E-07	7.10E-04	IGF2	-	-	-
10	2	rs81360254	9,737	0.446	9.97E-07	7.10E-04	IGF2	intron variant	ENSSSCG00000024015	FADS1
10	2	rs81343625	7,199	0.872	1.30E-06	8.91E-04	IGF2	-	-	-
10	2	rs81246704	7,324	0.128	1.30E-06	8.91E-04	IGF2	upstream gene variant	ENSSSCG00000013018	CDC42BPG
10	2	rs81359894	8,540	0.601	1.42E-06	9.60E-04	IGF2	-	-	-
10	2	rs81323907	12,604	0.725	1.61E-06	1.07E-03	IGF2	missense variant	ENSSSCG00000013157	OR5B21
10	2	rs81361056	10,099	0.603	1.71E-06	1.11E-03	IGF2	intron variant	ENSSSCG00000013083	CPSF7
10	2	rs81221368	6,457	0.749	1.86E-06	1.19E-03	IGF2	5 prime UTR variant	ENSSSCG00000012971	EFEMP2
10	2	rs81363153	13,192	0.625	1.96E-06	1.21E-03	IGF2	intron variant	ENSSSCG00000013174	CTNND1
10	2	rs81363209	13,218	0.375	1.96E-06	1.21E-03	IGF2	-	-	-
10	2	rs81362332	12,168	0.49	2.51E-06	1.53E-03	IGF2	-	-	-
10	2	rs338641431	9,757	0.695	2.61E-06	1.56E-03	IGF2	downstream gene variant	ENSSSCG00000039481	-
10	2	rs81315092	17,672	0.277	2.68E-06	1.56E-03	IGF2	upstream gene variant	ENSSSCG00000013278	TSPAN18
10	2	rs81368350	17,683	0.723	2.68E-06	1.56E-03	IGF2	-	-	-
10	2	rs81308303	9,495	0.249	2.96E-06	1.69E-03	IGF2	-	-	-
10	2	rs81333747	7,394	0.851	3.07E-06	1.74E-03	IGF2	intron variant	ENSSSCG00000013021	SF1
10	2	rs81363250	13,230	0.376	3.47E-06	1.93E-03	IGF2	3 prime UTR variant	ENSSSCG00000013176	-
10	2	rs81246105	10,892	0.479	4.02E-06	2.21E-03	IGF2	intron variant	ENSSSCG00000013106	PTGDR2
10	2	rs81362978	13,105	0.379	4.50E-06	2.44E-03	IGF2	-	-	-
10	2	rs81363413	13,307	0.36	5.17E-06	2.76E-03	IGF2	-	-	-
10	2	rs81358774	7,090	0.845	7.05E-06	3.71E-03	IGF2	intron variant	ENSSSCG00000013005	VPS51
10	2	rs81336616	11,627	0.431	9.10E-06	4.73E-03	IGF2	-	-	-
10	2	rs81474400	9,435	0.734	9.69E-06	4.97E-03	IGF2	intron variant	ENSSSCG00000013066	INCENP
10	2	rs81360403	9,588	0.754	1.21E-05	6.14E-03	IGF2	intron variant	ENSSSCG00000013074	RAB3IL1
10	2	rs318322737	13,167	0.523	1.44E-05	7.21E-03	IGF2	synonymous variant	ENSSSCG00000013174	CTNND1
10	2	rs330591156	11,057	0.159	1.50E-05	7.38E-03	IGF2	downstream gene variant	ENSSSCG00000031637	-
10	2	rs81262060	12,997	0.634	1.57E-05	7.61E-03	IGF2	upstream gene variant	ENSSSCG00000013170	-
10	2	rs81356796	5,827	0.208	1.73E-05	8.30E-03	IGF2	intron variant	ENSSSCG00000012934	CCS
10	2	rs81361375	10,584	0.324	1.87E-05	8.87E-03	IGF2	upstream gene variant	ENSSSCG00000029938	-
10	2	rs81341464	11,873	0.187	2.14E-05	9.77E-03	IGF2	upstream gene variant	ENSSSCG00000018435	RF00026
10	2	rs81474697	11,683	0.577	2.16E-05	9.77E-03	IGF2	intron variant	ENSSSCG00000013119	STX3
10	2	rs81253085	11,693	0.577	2.16E-05	9.77E-03	IGF2	intron variant	ENSSSCG00000013118	MRPL16
10	2	rs81322356	11,716	0.577	2.16E-05	9.77E-03	IGF2	intron variant	ENSSSCG00000013119	STX3
10	2	rs81368610	4,930	0.268	2.64E-05	1.18E-02	IGF2	downstream gene variant	ENSSSCG00000026349	ALDH3B2
10	2	rs81366508	16,231	0.48	2.95E-05	1.30E-02	IGF2	-	-	-
10	2	rs81357066	6,129	0.231	3.04E-05	1.33E-02	IGF2	upstream gene variant	ENSSSCG00000012955	KLC2
10	2	rs325325237	9,419	0.668	3.17E-05	1.37E-02	IGF2	intron variant	ENSSSCG00000013066	INCENP
10	2	rs81357655	6,591	0.563	4.20E-05	1.79E-02	IGF2	upstream gene variant	ENSSSCG00000012981	RELA
10	2	rs81331133	12,718	0.369	4.46E-05	1.86E-02	IGF2	upstream gene variant	ENSSSCG00000039026	-
10	2	rs81305360	12,725	0.631	4.46E-05	1.86E-02	IGF2	downstream gene variant	ENSSSCG00000039026	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
10	2	rs81357433	6,614	0.335	4.53E-05	1.87E-02	IGF2	downstream gene variant	ENSSSCG00000012983	PCNX3
10	2	rs81304212	25,964	0.507	4.88E-05	2.00E-02	IGF2	-	-	-
10	2	rs81313353	145,281	0.834	5.35E-05	2.16E-02	IGF2	-	-	-
10	2	rs81368115	16,738	0.728	5.97E-05	2.39E-02	IGF2	-	-	-
10	2	rs81270678	8,882	0.31	6.03E-05	2.39E-02	IGF2	-	-	-
10	2	rs81356578	5,840	0.415	7.62E-05	2.99E-02	IGF2	downstream gene variant	ENSSSCG00000012933	CCDC87
10	2	rs81362098	10,626	0.176	8.22E-05	3.19E-02	IGF2	intron variant	ENSSSCG00000028762	VPS37C
10	2	rs81341296	14,499	0.326	8.59E-05	3.30E-02	IGF2	-	-	-
10	2	rs81212188	7,829	0.689	8.80E-05	3.35E-02	IGF2	synonymous variant	ENSSSCG00000013032	GPR137
10	2	rs81362768	12,935	0.48	9.54E-05	3.59E-02	IGF2	-	-	-
10	2	rs81343787	7,363	0.648	1.01E-04	3.78E-02	IGF2	intron variant	ENSSSCG00000013019	MEN1
10	2	rs81360547	9,787	0.732	1.21E-04	4.44E-02	IGF2	intron variant	ENSSSCG00000013078	MYRF

Table S2: List of significant associated SNPs within eQTLs intervals for the 45-muscle gene expression study in each backcross independently.

BC1_LD:

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
1	1	rs325958068	135,745	0.022	5.61E-06	3.58E-03	ACSM5	intron variant	ENSSSCG00000035254	DPH6
1	1	rs81348596	135,974	0.978	5.61E-06	3.58E-03	ACSM5	non coding transcript variant	ENSSSCG00000033875	-
1	1	rs80962835	139,353	0.877	8.18E-06	4.72E-03	ACSM5	-	-	-
2	1	rs80863919	251,939	0.022	5.61E-06	3.58E-03	ACSM5	intron variant	ENSSSCG00000005457	-
2	1	rs81304807	254,874	0.925	5.94E-06	3.73E-03	ACSM5	intron variant	ENSSSCG00000005487	COL27A1
2	1	rs81352111	258,564	0.319	1.39E-05	7.31E-03	ACSM5	-	-	-
3	3	rs81308147	16,219	0.741	7.07E-05	2.85E-02	ACSM5	-	-	-
3	3	rs81339855	18,040	0.224	9.75E-05	3.62E-02	ACSM5	upstream gene variant	ENSSSCG00000025321	MAZ
3	3	rs81233426	18,505	0.158	1.39E-04	4.97E-02	ACSM5	upstream gene variant	ENSSSCG00000021845	ATXN2L
3	3	rs81322563	18,557	0.904	3.91E-07	4.18E-04	ACSM5	intron variant	ENSSSCG00000007805	ATP2A1
3	3	rs81477337	18,736	0.079	1.31E-05	7.02E-03	ACSM5	-	-	-
3	3	rs81316050	18,855	0.921	1.31E-05	7.02E-03	ACSM5	intron variant	ENSSSCG00000007812	XPO6
3	3	rs81378895	20,747	0.101	8.56E-07	7.59E-04	ACSM5	-	-	-
3	3	rs81378910	20,961	0.768	3.98E-06	2.89E-03	ACSM5	-	-	-
3	3	rs80838414	21,180	0.539	8.38E-06	4.76E-03	ACSM5	-	-	-
3	3	rs81321987	21,418	0.627	4.25E-06	3.03E-03	ACSM5	-	-	-
3	3	rs81379094	22,269	0.662	7.10E-06	4.30E-03	ACSM5	intron variant	ENSSSCG00000033674	PRKCB
3	3	rs81299374	22,286	0.338	7.10E-06	4.30E-03	ACSM5	intron variant	ENSSSCG00000033674	PRKCB
3	3	rs81477531	22,623	0.474	6.11E-07	6.00E-04	ACSM5	intron variant	ENSSSCG00000010795	-
3	3	rs81284839	22,634	0.526	6.11E-07	6.00E-04	ACSM5	downstream gene variant	ENSSSCG00000010795	-
3	3	rs81309807	22,735	0.303	1.08E-05	5.96E-03	ACSM5	downstream gene variant	ENSSSCG00000010798	GGA2
3	3	rs81309174	22,744	0.763	2.26E-06	1.75E-03	ACSM5	intron variant	ENSSSCG00000010799	COG7
3	3	rs81306471	22,905	0.268	3.21E-08	4.03E-05	ACSM5	upstream gene variant	ENSSSCG00000028923	SCNN1B

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
3	3	rs81319066	22,975	0.732	3.21E-08	4.03E-05	ACSM5	downstream gene variant	ENSSSCG00000007836	SCNN1G
3	3	rs81330380	23,056	0.735	1.74E-08	3.01E-05	ACSM5	intron variant	ENSSSCG00000030424	USP31
3	3	rs80876065	23,093	0.268	3.21E-08	4.03E-05	ACSM5	downstream gene variant	ENSSSCG00000023632	RF00414
3	3	rs81321464	23,166	0.395	7.54E-05	2.90E-02	ACSM5	-	-	-
3	3	rs81312070	23,335	0.801	2.59E-09	9.43E-06	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
3	3	rs81227560	23,359	0.202	1.45E-09	5.85E-06	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
3	3	rs81238947	23,519	0.268	3.21E-08	4.03E-05	ACSM5	-	-	-
3	3	rs81324887	23,568	0.754	4.03E-07	4.19E-04	ACSM5	downstream gene variant	ENSSSCG00000007838	OTOA
3	3	rs81340946	23,582	0.197	5.61E-06	3.58E-03	ACSM5	intron variant	ENSSSCG00000007838	OTOA
3	3	rs81344302	23,621	0.461	6.63E-05	2.74E-02	ACSM5	intron variant	ENSSSCG00000007838	OTOA
3	3	rs81314080	23,755	0.333	5.20E-05	2.20E-02	ACSM5	synonymous variant	ENSSSCG00000036268	-
3	3	rs81344425	23,814	0.592	8.05E-06	4.72E-03	ACSM5	intron variant	ENSSSCG00000032340	-
3	3	rs81335819	23,818	0.316	2.01E-05	1.02E-02	ACSM5	intron variant	ENSSSCG00000032340	-
3	3	rs81313849	23,887	0.711	1.49E-06	1.18E-03	ACSM5	upstream gene variant	ENSSSCG00000020312	RF00026
3	3	rs81308074	23,896	0.289	1.49E-06	1.18E-03	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
3	3	rs323881880	23,900	0.289	1.49E-06	1.18E-03	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
3	3	rs81322298	23,986	0.224	1.77E-11	1.08E-07	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
3	3	rs81379135	24,070	0.469	3.87E-05	1.68E-02	ACSM5	intron variant	ENSSSCG00000025266	VWA3A
3	3	rs81336511	24,365	0.627	9.10E-05	3.45E-02	ACSM5	intron variant	ENSSSCG00000035000	-
3	3	rs81234875	24,529	0.846	1.57E-08	3.01E-05	ACSM5	-	-	-
3	3	rs81278892	24,815	0.197	5.61E-06	3.58E-03	ACSM5	5 prime UTR variant	ENSSSCG00000007849	CRYM
3	3	rs81326798	24,892	0.732	3.21E-08	4.03E-05	ACSM5	intron variant	ENSSSCG00000031876	TMEM159
3	3	rs81336877	24,920	0.675	3.74E-07	4.13E-04	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
3	3	rs81288253	24,942	0.719	1.78E-08	3.01E-05	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
3	3	rs81315383	24,980	0.765	4.92E-06	3.44E-03	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
3	3	rs81249771	25,004	0.675	7.99E-06	4.72E-03	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
3	3	rs81379171	25,072	0.741	1.76E-08	3.01E-05	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
3	3	rs81379199	25,155	0.728	2.28E-08	3.45E-05	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
3	3	rs81379197	25,173	0.272	2.28E-08	3.45E-05	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
3	3	rs81315362	25,206	0.741	4.87E-09	1.18E-05	ACSM5	intron variant	ENSSSCG00000007855	REXO5
3	3	rs81288413	25,217	0.259	4.87E-09	1.18E-05	ACSM5	intron variant	ENSSSCG00000007855	REXO5
3	3	rs81335959	25,260	0.259	4.87E-09	1.18E-05	ACSM5	intron variant	ENSSSCG00000007857	ACSM3
3	3	rs81324695	25,404	0.662	1.32E-04	4.79E-02	ACSM5	intron variant	ENSSSCG00000026453	ACSM5
3	3	ACSM5.P	25,422	0.216	3.22E-13	1.17E-08	ACSM5	upstream gene variant	ENSSSCG00000026453	ACSM5
3	3	rs81326933	25,540	0.763	6.93E-08	8.40E-05	ACSM5	-	-	-
3	3	rs81347321	25,606	0.289	1.55E-05	8.06E-03	ACSM5	-	-	-
3	3	rs81278505	25,651	0.167	1.24E-11	8.99E-08	ACSM5	-	-	-
3	3	rs81240993	25,654	0.211	1.03E-06	8.96E-04	ACSM5	-	-	-
3	3	rs81244431	25,758	0.263	1.66E-08	3.01E-05	ACSM5	downstream gene variant	ENSSSCG00000007862	GPR139
3	3	rs81247258	25,965	0.276	1.82E-08	3.01E-05	ACSM5	intron variant	ENSSSCG00000034720	IQCK
3	3	rs333552464	26,183	0.724	1.82E-08	3.01E-05	ACSM5	intron variant	ENSSSCG00000023829	CCP110
3	3	rs81333672	26,196	0.57	2.65E-05	1.20E-02	ACSM5	intron variant	ENSSSCG00000029212	GDE1
3	3	rs81475002	26,275	0.772	1.76E-12	1.60E-08	ACSM5	intron variant	ENSSSCG00000007868	TMC5
3	3	rs81329230	26,346	0.522	2.46E-06	1.86E-03	ACSM5	intron variant	ENSSSCG00000007866	TMC7
3	3	rs324741666	26,460	0.772	1.76E-12	1.60E-08	ACSM5	intron variant	ENSSSCG00000022200	SYT17
3	3	rs81474976	26,514	0.772	1.76E-12	1.60E-08	ACSM5	intron variant	ENSSSCG00000022200	SYT17
3	3	rs81475285	26,619	0.19	4.40E-05	1.88E-02	ACSM5	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
3	3	rs81239835	26,750	0.789	8.37E-10	3.80E-06	ACSM5	intron variant	ENSSSCG00000032531	SMG1
3	3	rs81267534	26,970	0.443	1.97E-05	1.01E-02	ACSM5	-	-	-
3	3	rs81379308	27,015	0.452	8.18E-07	7.44E-04	ACSM5	-	-	-
3	3	rs81379367	27,201	0.526	2.88E-05	1.28E-02	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
3	3	rs81379369	27,217	0.289	1.26E-07	1.43E-04	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
3	3	rs81238437	27,259	0.205	5.15E-10	2.68E-06	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
3	3	rs81379421	27,307	0.341	3.25E-05	1.43E-02	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
3	3	rs81379431	27,322	0.289	1.26E-07	1.43E-04	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
3	3	rs81319789	33,568	0.25	1.44E-06	1.18E-03	ACSM5	-	-	-
3	3	rs81334729	34,824	0.689	6.25E-05	2.61E-02	ACSM5	-	-	-
3	3	rs81478928	36,083	0.763	9.87E-06	5.52E-03	ACSM5	-	-	-
3	3	rs81289409	36,102	0.224	2.44E-05	1.18E-02	ACSM5	-	-	-
3	3	rs81266926	45,069	0.246	2.54E-05	1.18E-02	ACSM5	-	-	-
3	3	rs81312320	45,096	0.754	2.54E-05	1.18E-02	ACSM5	-	-	-
3	3	rs81224450	45,151	0.754	2.54E-05	1.18E-02	ACSM5	-	-	-
3	3	rs81237540	48,604	0.154	2.53E-05	1.18E-02	ACSM5	-	-	-
4	6	rs81336707	20,167	0.022	7.72E-07	7.20E-04	ACSM5	intron variant	ENSSSCG00000002802	GINS3
4	6	rs81336454	26,053	0.978	7.72E-07	7.20E-04	ACSM5	-	-	-
4	6	rs81395172	32,407	0.026	2.11E-05	1.05E-02	ACSM5	-	-	-
5	8	rs81343181	19,994	0.184	1.42E-04	5.00E-02	ACSM5	intron variant	ENSSSCG00000022155	RBPJ
5	8	rs81330366	20,011	0.123	2.68E-05	1.20E-02	ACSM5	intron variant	ENSSSCG00000022155	RBPJ
5	8	rs81476978	20,038	0.123	2.68E-05	1.20E-02	ACSM5	intron variant	ENSSSCG00000022155	RBPJ
5	8	rs81406761	20,481	0.996	3.50E-09	1.06E-05	ACSM5	intron variant	ENSSSCG00000008762	STIM2
6	10	rs81252142	60,180	0.009	7.30E-05	2.85E-02	ACSM5	-	-	-
6	10	rs81426812	60,532	0.991	7.30E-05	2.85E-02	ACSM5	intron variant	ENSSSCG00000011121	-
6	10	rs81260789	60,574	0.009	7.30E-05	2.85E-02	ACSM5	intron variant	ENSSSCG00000011121	-
6	10	rs81253334	60,634	0.009	7.30E-05	2.85E-02	ACSM5	intron variant	ENSSSCG00000011121	-
7	2	rs81284541	119,937	0.004	1.91E-07	4.34E-04	HIF1AN	intron variant	ENSSSCG00000014219	CDO1
7	2	rs81363852	120,690	0.004	1.91E-07	4.34E-04	HIF1AN	upstream gene variant	ENSSSCG00000014224	SEMA6A
7	2	rs81363933	122,105	0.996	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81363986	122,538	0.004	2.08E-07	4.44E-04	HIF1AN	-	-	-
7	2	rs81225815	122,722	0.004	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81364080	123,299	0.996	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81364093	123,395	0.004	1.91E-07	4.34E-04	HIF1AN	intron variant	ENSSSCG00000028431	HSD17B4
7	2	rs81295472	123,998	0.996	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81364195	124,059	0.004	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81296107	124,163	0.996	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs80790446	124,860	0.004	1.91E-07	4.34E-04	HIF1AN	-	-	-
7	2	rs81474819	128,693	0.004	1.91E-07	4.34E-04	HIF1AN	-	-	-
8	5	rs81386076	86,771	0.013	3.10E-06	5.63E-03	HIF1AN	-	-	-
8	5	rs81337794	92,956	0.982	2.49E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000027898	ATP2B1
8	5	rs81311166	93,003	0.982	2.49E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000027898	ATP2B1
9	7	rs80870930	62,413	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80824617	62,760	0.009	4.09E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000039714	MIPOL1
9	7	rs80949277	63,306	0.991	4.09E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000039317	SLC25A21
9	7	rs80866076	63,442	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs321616390	63,701	0.991	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80975688	64,175	0.991	4.09E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000032377	RALGAPA1
9	7	rs80971610	64,386	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80979456	64,459	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80914087	64,482	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80921778	64,607	0.991	4.09E-05	3.72E-02	HIF1AN	upstream gene variant	ENSSSCG00000001951	PSMA6

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
9	7	rs80803727	64,812	0.009	4.09E-05	3.72E-02	HIF1AN	upstream gene variant	ENSSSCG00000034753	FAM177A1
9	7	rs80911625	65,210	0.991	4.09E-05	3.72E-02	HIF1AN	upstream gene variant	ENSSSCG00000032041	RF00003
9	7	rs80933492	65,815	0.991	4.09E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000001963	EGLN3
9	7	rs80995643	65,863	0.991	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80788814	65,883	0.009	4.09E-05	3.72E-02	HIF1AN	-	-	-
9	7	rs80961115	65,963	0.009	4.09E-05	3.72E-02	HIF1AN	intron variant	ENSSSCG00000001964	NPAS3
10	2	rs81284541	119,937	0.004	3.24E-05	3.20E-02	ChREBP	intron variant	ENSSSCG00000014219	CDO1
10	2	rs81363852	120,690	0.004	3.24E-05	3.20E-02	ChREBP	upstream gene variant	ENSSSCG00000014224	SEMA6A
10	2	rs81363933	122,105	0.996	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81363986	122,538	0.004	3.31E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81225815	122,722	0.004	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81364080	123,299	0.996	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81364093	123,395	0.004	3.24E-05	3.20E-02	ChREBP	intron variant	ENSSSCG00000028431	HSD17B4
10	2	rs81295472	123,998	0.996	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81364195	124,059	0.004	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81296107	124,163	0.996	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs80790446	124,860	0.004	3.24E-05	3.20E-02	ChREBP	-	-	-
10	2	rs81474819	128,693	0.004	3.24E-05	3.20E-02	ChREBP	-	-	-
11	9	rs81420563	18,204	0.974	4.75E-07	2.47E-03	ChREBP	intron variant	ENSSSCG00000014904	DLG2
11	9	rs81281297	18,261	0.026	4.75E-07	2.47E-03	ChREBP	intron variant	ENSSSCG00000014904	DLG2
11	9	rs81407294	18,855	0.026	4.75E-07	2.47E-03	ChREBP	-	-	-
11	9	rs81407308	18,952	0.026	4.75E-07	2.47E-03	ChREBP	-	-	-
11	9	rs81407558	21,105	0.026	4.75E-07	2.47E-03	ChREBP	-	-	-
11	9	rs81408056	25,964	0.86	1.81E-05	3.20E-02	ChREBP	-	-	-
11	9	rs81408162	26,349	0.86	1.81E-05	3.20E-02	ChREBP	-	-	-
11	9	rs81408173	26,457	0.75	6.11E-05	5.05E-02	ChREBP	intron variant	ENSSSCG00000032360	PANX1
12	13	rs81447133	81,152	0.399	6.95E-06	2.53E-02	ChREBP	-	-	-
12	13	rs81447169	81,254	0.281	2.91E-05	3.20E-02	ChREBP	-	-	-
12	13	rs322937606	81,420	0.702	2.48E-05	3.20E-02	ChREBP	intron variant	ENSSSCG00000011666	CLSTN2
12	13	rs80946349	81,440	0.702	2.48E-05	3.20E-02	ChREBP	intron variant	ENSSSCG00000011666	CLSTN2
12	13	rs80916261	81,488	0.298	2.48E-05	3.20E-02	ChREBP	-	-	-
12	13	rs80942072	81,561	0.298	2.48E-05	3.20E-02	ChREBP	intron variant	ENSSSCG00000011668	TRIM42
12	13	rs81316562	81,604	0.298	2.48E-05	3.20E-02	ChREBP	-	-	-
12	13	rs81344961	81,639	0.342	3.69E-05	3.44E-02	ChREBP	-	-	-
13	2	rs81284541	119,937	0.004	6.42E-11	1.46E-07	CREG1	intron variant	ENSSSCG00000014219	CDO1
13	2	rs81363852	120,690	0.004	6.42E-11	1.46E-07	CREG1	upstream gene variant	ENSSSCG00000014224	SEMA6A
13	2	rs81363933	122,105	0.996	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81363986	122,538	0.004	6.35E-11	1.46E-07	CREG1	-	-	-
13	2	rs81225815	122,722	0.004	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81364080	123,299	0.996	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81364093	123,395	0.004	6.42E-11	1.46E-07	CREG1	intron variant	ENSSSCG00000028431	HSD17B4
13	2	rs81295472	123,998	0.996	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81364195	124,059	0.004	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81296107	124,163	0.996	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs80790446	124,860	0.004	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81474819	128,693	0.004	6.42E-11	1.46E-07	CREG1	-	-	-
13	2	rs81365226	133,980	0.991	2.08E-06	3.03E-03	CREG1	intron variant	ENSSSCG00000014269	FNIP1
13	2	rs81326721	135,397	0.991	2.08E-06	3.03E-03	CREG1	downstream gene variant	ENSSSCG00000014292	HSPA4
13	2	rs81365594	135,551	0.009	2.08E-06	3.03E-03	CREG1	intron variant	ENSSSCG00000014293	FSTL4
Inter-	**Chr**	**SNP**	**Position**	**MAF**	**p-value**	**FDR**	**Associated**	**Consequence**	**Ensembl Gene ID**	**Gene**

val			(Mb)				Gene			Symbol
13	2	rs81365661	135,636	0.009	2.08E-06	3.03E-03	CREG1	intron variant	ENSSSCG00000014293	FSTL4
13	2	rs81295831	135,766	0.009	2.08E-06	3.03E-03	CREG1	intron variant	ENSSSCG00000014293	FSTL4
13	2	rs81338611	136,199	0.009	2.08E-06	3.03E-03	CREG1	-	-	-
14	2	rs81284541	119,937	0.004	4.27E-08	9.14E-05	DGAT2	intron variant	ENSSSCG00000014219	CDO1
14	2	rs81363852	120,690	0.004	4.27E-08	9.14E-05	DGAT2	upstream gene variant	ENSSSCG00000014224	SEMA6A
14	2	rs81363933	122,105	0.996	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81363986	122,538	0.004	4.26E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81225815	122,722	0.004	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81364080	123,299	0.996	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81364093	123,395	0.004	4.27E-08	9.14E-05	DGAT2	intron variant	ENSSSCG00000028431	HSD17B4
14	2	rs81295472	123,998	0.996	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81364195	124,059	0.004	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81296107	124,163	0.996	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs80790446	124,860	0.004	4.27E-08	9.14E-05	DGAT2	-	-	-
14	2	rs81474819	128,693	0.004	4.27E-08	9.14E-05	DGAT2	-	-	-
15	7	rs80870930	62,413	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80824617	62,760	0.009	3.88E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000039714	MIPOL1
15	7	rs80949277	63,306	0.991	3.88E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000039317	SLC25A21
15	7	rs80866076	63,442	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs321616390	63,701	0.991	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80975688	64,175	0.991	3.88E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000032377	RALGAPA1
15	7	rs80971610	64,386	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80979456	64,459	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80914087	64,482	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80921778	64,607	0.991	3.88E-05	3.53E-02	DGAT2	upstream gene variant	ENSSSCG00000001951	PSMA6
15	7	rs80803727	64,812	0.009	3.88E-05	3.53E-02	DGAT2	upstream gene variant	ENSSSCG00000034753	FAM177A1
15	7	rs80911625	65,210	0.991	3.88E-05	3.53E-02	DGAT2	upstream gene variant	ENSSSCG00000032041	RF00003
15	7	rs80933492	65,815	0.991	3.88E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000001963	EGLN3
15	7	rs80835896	65,827	0.039	2.77E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000001963	EGLN3
15	7	rs80995643	65,863	0.991	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80788814	65,883	0.009	3.88E-05	3.53E-02	DGAT2	-	-	-
15	7	rs80961115	65,963	0.009	3.88E-05	3.53E-02	DGAT2	intron variant	ENSSSCG00000001964	NPAS3
16	9	rs81415303	106,297	0.987	6.33E-05	4.80E-02	DGAT2	-	-	-
16	9	rs81415878	112,941	0.013	6.33E-05	4.80E-02	DGAT2	intron variant	ENSSSCG00000015462	TPK1
16	9	rs81300533	112,954	0.987	6.33E-05	4.80E-02	DGAT2	intron variant	ENSSSCG00000015462	TPK1
16	9	rs80886851	112,974	0.013	6.33E-05	4.80E-02	DGAT2	intron variant	ENSSSCG00000015462	TPK1
16	9	rs81415886	113,055	0.013	6.33E-05	4.80E-02	DGAT2	intron variant	ENSSSCG00000015462	TPK1
16	9	rs81416742	123,003	0.013	6.33E-05	4.80E-02	DGAT2	intron variant	ENSSSCG00000015543	CACNA1E
17	11	rs80796231	8,855	0.128	8.29E-07	1.00E-02	FOS	intron variant	ENSSSCG00000029039	BRCA2
17	11	rs80845358	10,367	0.877	3.72E-09	1.35E-04	FOS	-	-	-
17	11	rs80798788	10,399	0.882	2.44E-08	4.45E-04	FOS	-	-	-
18	2	rs81306755	145	0.509	4.43E-07	1.07E-03	IGF2	intron variant	ENSSSCG00000014565	-
18	2	rs81328276	236	0.246	1.83E-09	8.34E-06	IGF2	intron variant	ENSSSCG00000024569	ANO9
18	2	rs81317307	310	0.654	5.66E-06	9.80E-03	IGF2	intron variant	ENSSSCG00000027045	LRRC56
18	2	rs81339115	422	0.325	4.26E-06	7.74E-03	IGF2	intron variant	ENSSSCG00000012850	DEAF1
18	2	rs81341763	677	0.254	4.44E-10	2.31E-06	IGF2	-	-	-
18	2	IGF2	1,483	0.794	3.03E-07	7.87E-04	IGF2	-	-	-
18	2	rs81252426	2,984	0.746	2.34E-07	6.56E-04	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81330112	3,058	0.746	2.34E-07	6.56E-04	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81336288	3,062	0.811	1.20E-10	1.50E-06	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81341267	3,094	0.746	2.34E-07	6.56E-04	IGF2	intron variant	ENSSSCG00000031191	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
18	2	rs81328266	3,657	0.663	8.89E-07	1.80E-03	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81322199	3,689	0.075	1.45E-15	5.29E-11	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81356987	3,859	0.693	1.98E-10	1.64E-06	IGF2	-	-	-
18	2	rs81368353	3,895	0.382	7.63E-06	1.11E-02	IGF2	-	-	-
18	2	rs81364067	4,412	0.465	1.84E-05	2.31E-02	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
18	2	rs81364734	4,444	0.478	1.85E-06	3.54E-03	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
18	2	rs81337384	4,531	0.636	8.42E-07	1.80E-03	IGF2	downstream gene variant	ENSSSCG00000012885	-
18	2	rs81368683	4,966	0.197	1.24E-10	1.50E-06	IGF2	upstream gene variant	ENSSSCG00000012896	NDUFV1
18	2	rs81355859	5,069	0.794	1.70E-08	6.87E-05	IGF2	intron variant	ENSSSCG00000012906	CABP4
18	2	rs81326091	5,417	0.259	7.50E-06	1.11E-02	IGF2	intron variant	ENSSSCG00000024837	SYT12
18	2	rs81356888	5,928	0.781	3.62E-05	3.88E-02	IGF2	intron variant	ENSSSCG00000030415	DPP3
18	2	rs81357012	6,072	0.798	2.25E-10	1.64E-06	IGF2	missense variant	ENSSSCG00000029949	CD248
18	2	rs81359193	7,493	0.228	7.34E-06	1.11E-02	IGF2	intron variant	ENSSSCG00000033188	-
18	2	rs81359004	7,536	0.154	2.71E-10	1.64E-06	IGF2	intron variant	ENSSSCG00000013024	-
18	2	rs81278072	7,642	0.272	7.34E-06	1.11E-02	IGF2	-	-	-
18	2	rs81316921	8,919	0.224	1.63E-05	2.12E-02	IGF2	upstream gene variant	ENSSSCG00000026837	RF00087
18	2	rs81474654	9,224	0.763	2.18E-07	6.56E-04	IGF2	intron variant	ENSSSCG00000036669	-
18	2	rs81474931	11,764	0.136	6.15E-07	1.40E-03	IGF2	non coding transcript exon variant	ENSSSCG00000037095	-
19	9	rs81277796	54,051	0.272	1.14E-04	3.21E-02	MGLL	-	-	-
19	9	rs81314908	54,054	0.298	2.38E-05	8.33E-03	MGLL	-	-	-
19	9	rs81411620	55,054	0.355	1.29E-04	3.51E-02	MGLL	non coding transcript variant	ENSSSCG00000035237	-
19	9	rs81411623	55,068	0.355	1.29E-04	3.51E-02	MGLL	non coding transcript variant	ENSSSCG00000035237	-
19	9	rs81276455	55,489	0.25	1.22E-04	3.39E-02	MGLL	intron variant	ENSSSCG00000015235	ETS1
19	9	rs81477388	55,857	0.202	9.67E-05	2.81E-02	MGLL	synonymous variant	ENSSSCG00000021573	KCNJ5
20	13	rs80813451	47,034	0.465	1.67E-04	4.39E-02	MGLL	intron variant	ENSSSCG00000011497	MAGI1
20	13	rs334822364	48,282	0.596	9.74E-05	2.81E-02	MGLL	-	-	-
20	13	rs80901843	50,392	0.627	6.05E-05	1.85E-02	MGLL	-	-	-
20	13	rs81213102	50,537	0.627	6.05E-05	1.85E-02	MGLL	synonymous variant	ENSSSCG00000033154	ARL6IP5
20	13	rs81445590	51,387	0.496	1.23E-06	1.41E-03	MGLL	intron variant	ENSSSCG00000011514	MITF
20	13	rs81445884	54,095	0.68	8.98E-06	4.31E-03	MGLL	intron variant	ENSSSCG00000011518	SHQ1
20	13	rs81274501	54,132	0.32	8.98E-06	4.31E-03	MGLL	intron variant	ENSSSCG00000011518	SHQ1
20	13	rs81445986	54,676	0.421	1.39E-04	3.74E-02	MGLL	intron variant	ENSSSCG00000011521	PDZRN3
20	13	rs80927938	54,846	0.61	9.59E-05	2.81E-02	MGLL	-	-	-
20	13	rs81446005	55,042	0.803	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81294606	55,361	0.211	1.32E-06	1.41E-03	MGLL	-	-	-
20	13	rs81246453	55,420	0.211	1.32E-06	1.41E-03	MGLL	-	-	-
20	13	rs80831322	55,876	0.211	1.32E-06	1.41E-03	MGLL	intron variant	ENSSSCG00000011522	CNTN3
20	13	rs81446067	55,930	0.797	1.07E-06	1.39E-03	MGLL	intron variant	ENSSSCG00000011522	CNTN3
20	13	rs81446075	55,997	0.197	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000011522	CNTN3
20	13	rs81478408	56,027	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs345179433	57,384	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81340615	57,486	0.789	1.32E-06	1.41E-03	MGLL	-	-	-
20	13	rs81446093	57,604	0.211	1.32E-06	1.41E-03	MGLL	-	-	-
20	13	rs80909709	58,821	0.39	9.59E-05	2.81E-02	MGLL	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
20	13	rs80995262	59,654	0.504	6.21E-05	1.88E-02	MGLL	intron variant	ENSSSCG00000011528	IL5RA
20	13	rs81344292	59,846	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs80827353	60,065	0.434	1.99E-04	5.04E-02	MGLL	-	-	-
20	13	rs80841140	60,113	0.434	1.99E-04	5.04E-02	MGLL	-	-	-
20	13	rs81446158	60,146	0.566	1.99E-04	5.04E-02	MGLL	-	-	-
20	13	rs81446170	60,259	0.478	2.79E-05	9.40E-03	MGLL	-	-	-
20	13	rs81223384	60,431	0.566	1.99E-04	5.04E-02	MGLL	downstream gene variant	ENSSSCG00000023806	LRRN1
20	13	rs81344912	60,744	0.803	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs80930641	60,906	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81274392	60,981	0.803	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000011532	SUMF1
20	13	rs81446188	60,995	0.197	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000011532	SUMF1
20	13	rs81446205	61,107	0.803	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000023437	ITPR1
20	13	rs81446211	61,140	0.478	2.00E-05	7.42E-03	MGLL	intron variant	ENSSSCG00000023437	ITPR1
20	13	rs81446213	61,184	0.197	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000023437	ITPR1
20	13	rs81446232	61,461	0.803	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81446235	61,518	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81298552	61,643	0.803	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000011535	ARL8B
20	13	rs81446263	61,718	0.803	2.68E-07	5.74E-04	MGLL	intron variant	ENSSSCG00000034197	EDEM1
20	13	rs81285927	61,779	0.803	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs80816799	61,862	0.197	2.68E-07	5.74E-04	MGLL	-	-	-
20	13	rs81294655	62,060	0.482	1.23E-04	3.39E-02	MGLL	-	-	-
20	13	rs81446290	62,144	0.596	1.60E-04	4.24E-02	MGLL	-	-	-
20	13	rs81298791	64,450	0.539	1.83E-05	6.86E-03	MGLL	-	-	-
20	13	rs81446362	64,483	0.447	1.39E-05	5.25E-03	MGLL	-	-	-
20	13	rs81446358	64,510	0.689	4.96E-07	8.08E-04	MGLL	-	-	-
20	13	rs81446433	64,996	0.509	1.12E-05	4.31E-03	MGLL	downstream gene variant	ENSSSCG00000011538	LMCD1
20	13	rs81241557	65,039	0.509	1.12E-05	4.31E-03	MGLL	3 prime UTR variant	ENSSSCG00000023082	SSUH2
20	13	rs81227824	65,597	0.478	6.67E-06	4.31E-03	MGLL	-	-	-
20	13	rs80971212	65,822	0.522	6.67E-06	4.31E-03	MGLL	intron variant	ENSSSCG00000011540	SETD5
20	13	rs80884874	65,882	0.522	6.67E-06	4.31E-03	MGLL	intron variant	ENSSSCG00000011543	LHFPL4
20	13	rs80953937	65,976	0.342	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011546	MTMR14
20	13	rs331955329	66,004	0.658	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011546	MTMR14
20	13	rs80971430	66,026	0.342	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011546	MTMR14
20	13	rs80895088	66,040	0.658	6.00E-07	8.08E-04	MGLL	downstream gene variant	ENSSSCG00000011546	MTMR14
20	13	rs80945527	66,104	0.342	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011553	-
20	13	rs80898778	66,130	0.623	1.18E-04	3.31E-02	MGLL	intron variant	ENSSSCG00000011555	RPUSD3
20	13	rs80885182	66,270	0.342	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011563	FANCD2
20	13	rs80932483	66,319	0.114	2.87E-05	9.40E-03	MGLL	intron variant	ENSSSCG00000011565	BRK1
20	13	rs81446446	66,388	0.658	6.00E-07	8.08E-04	MGLL	intron variant	ENSSSCG00000011570	IRAK2
20	13	rs45430493	66,515	0.658	6.00E-07	8.08E-04	MGLL	synonymous variant	ENSSSCG00000011572	SEC13
20	13	rs81446475	66,725	0.386	2.32E-05	8.18E-03	MGLL	intron variant	ENSSSCG00000023084	ATP2B2
20	13	rs81446455	66,763	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000023084	ATP2B2
20	13	rs81446484	66,777	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000023084	ATP2B2
20	13	rs81478601	66,795	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000023084	ATP2B2
20	13	rs81275610	66,799	0.693	4.46E-06	3.96E-03	MGLL	intron variant	ENSSSCG00000023084	ATP2B2
20	13	rs81312729	67,010	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000039850	SLC6A11
20	13	rs81291119	67,017	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000039850	SLC6A11
20	13	rs81227118	67,031	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000039850	SLC6A11
20	13	rs80786631	67,209	0.68	1.11E-05	4.31E-03	MGLL	5 prime UTR variant	ENSSSCG00000011576	HRH1

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
20	13	rs80887797	67,221	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011576	HRH1
20	13	rs80852045	67,240	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011576	HRH1
20	13	rs80841003	67,303	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011576	HRH1
20	13	rs80968667	67,313	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011576	HRH1
20	13	rs80956263	67,345	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011576	HRH1
20	13	rs81446503	67,460	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011575	ATG7
20	13	rs81291303	67,581	0.114	2.87E-05	9.40E-03	MGLL	intron variant	ENSSSCG00000011575	ATG7
20	13	rs81446559	67,849	0.307	4.46E-06	3.96E-03	MGLL	intron variant	ENSSSCG00000030581	VGLL4
20	13	rs81446534	67,869	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000030581	VGLL4
20	13	rs81286362	67,883	0.68	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81446547	67,956	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011578	TAMM41
20	13	rs81446550	68,050	0.68	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81446551	68,063	0.32	1.11E-05	4.31E-03	MGLL	upstream gene variant	ENSSSCG00000019373	RF00002
20	13	rs81446577	68,325	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011579	PPARG
20	13	rs81219146	68,363	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011579	PPARG
20	13	rs81446594	68,450	0.886	2.87E-05	9.40E-03	MGLL	-	-	-
20	13	rs81298754	68,477	0.114	2.87E-05	9.40E-03	MGLL	intron variant	ENSSSCG00000011580	TSEN2
20	13	rs81293162	68,507	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011580	TSEN2
20	13	rs81252658	68,804	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011587	EFCAB12
20	13	rs333600445	68,909	0.68	1.11E-05	4.31E-03	MGLL	downstream gene variant	ENSSSCG00000011589	IFT122
20	13	rs80970440	68,939	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011591	H1FOO
20	13	rs81446603	68,979	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011592	PLXND1
20	13	rs80982389	68,998	0.32	1.11E-05	4.31E-03	MGLL	upstream gene variant	ENSSSCG00000011592	PLXND1
20	13	rs80884918	69,011	0.321	2.10E-06	2.18E-03	MGLL	-	-	-
20	13	rs80921344	69,035	0.32	1.11E-05	4.31E-03	MGLL	downstream gene variant	ENSSSCG00000011593	TMCC1
20	13	rs80869386	69,052	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011593	TMCC1
20	13	rs81222920	69,940	0.474	1.55E-04	4.14E-02	MGLL	-	-	-
20	13	rs81446640	70,122	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011600	SLC6A6
20	13	rs81446653	70,211	0.32	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81446680	70,435	0.649	5.33E-05	1.66E-02	MGLL	-	-	-
20	13	rs81306386	70,558	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000039010	WNT7A
20	13	rs81327543	70,784	0.614	2.32E-05	8.18E-03	MGLL	intron variant	ENSSSCG00000011609	FBLN2
20	13	rs81345179	71,068	0.886	2.87E-05	9.40E-03	MGLL	intron variant	ENSSSCG00000027812	IQSEC1
20	13	rs81256676	71,082	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000027812	IQSEC1
20	13	rs81283643	71,375	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000023354	ACAD9
20	13	rs81315570	71,447	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000024552	-
20	13	rs81339681	71,513	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000024000	EFCC1
20	13	rs81215583	71,905	0.68	1.11E-05	4.31E-03	MGLL	synonymous variant	ENSSSCG00000011612	RPN1
20	13	rs81244930	73,122	0.351	3.72E-06	3.47E-03	MGLL	intron variant	ENSSSCG00000031557	CHCHD6
20	13	rs81312560	73,216	0.315	3.94E-07	7.96E-04	MGLL	intron variant	ENSSSCG00000031557	CHCHD6
20	13	rs80838457	73,367	0.649	3.72E-06	3.47E-03	MGLL	upstream gene variant	ENSSSCG00000011625	CHST13
20	13	rs80808704	73,510	0.412	1.02E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011627	ACPP
20	13	rs80985385	73,790	0.351	3.72E-06	3.47E-03	MGLL	intron variant	ENSSSCG00000011629	ACAD11
20	13	rs334940347	73,819	0.351	3.72E-06	3.47E-03	MGLL	upstream gene variant	ENSSSCG00000011631	UBA5
20	13	rs81478496	74,266	0.346	4.45E-05	1.39E-02	MGLL	-	-	-
20	13	rs81247562	74,363	0.68	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81239673	74,376	0.32	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81235136	74,455	0.32	1.11E-05	4.31E-03	MGLL	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
20	13	rs80804638	74,539	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011635	TMEM108
20	13	rs80869440	74,621	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011636	BFSP2
20	13	rs81220271	74,952	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011640	TF
20	13	rs81298520	74,997	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000011639	SRPRB
20	13	rs81446767	75,051	0.68	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000036503	RAB6B
20	13	rs81446765	75,073	0.32	1.11E-05	4.31E-03	MGLL	intron variant	ENSSSCG00000036503	RAB6B
20	13	rs81446763	75,090	0.68	1.11E-05	4.31E-03	MGLL	-	-	-
20	13	rs81446856	75,744	0.886	2.87E-05	9.40E-03	MGLL	intron variant	ENSSSCG00000011646	KY
20	13	rs81447107	80,895	0.237	2.20E-05	7.91E-03	MGLL	-	-	-
20	13	rs81447112	80,937	0.237	2.20E-05	7.91E-03	MGLL	-	-	-
21	13	rs80988778	107,655	0.461	2.95E-05	9.50E-03	MGLL	intron variant	ENSSSCG00000035525	-
21	13	rs80874675	107,736	0.461	2.95E-05	9.50E-03	MGLL	-	-	-
21	13	rs318579958	109,633	0.939	1.95E-04	5.04E-02	MGLL	-	-	-
21	13	rs81242841	109,659	0.939	1.95E-04	5.04E-02	MGLL	intron variant	ENSSSCG00000028905	TNIK
22	4	rs80803396	64,491	0.316	2.32E-06	4.26E-02	NCOA2	-	-	-
22	4	rs80949198	64,835	0.254	3.51E-06	4.26E-02	NCOA2	intron variant	ENSSSCG00000006193	TRAM1
22	4	rs80951880	65,609	0.268	3.43E-06	4.26E-02	NCOA2	intron variant	ENSSSCG00000006197	SULF1
22	4	rs81334801	69,900	0.25	6.51E-06	4.73E-02	NCOA2	-	-	-
23	16	rs81461074	63,606	0.215	3.23E-06	2.35E-02	PIK3R1	upstream gene variant	ENSSSCG00000017041	ADRA1B
23	16	rs81461119	63,755	0.785	3.23E-06	2.35E-02	PIK3R1	-	-	-
23	16	rs81222211	63,825	0.785	3.23E-06	2.35E-02	PIK3R1	-	-	-
24	2	rs81284541	119,937	0.004	1.44E-11	3.10E-08	PPARG	intron variant	ENSSSCG00000014219	CDO1
24	2	rs81363852	120,690	0.004	1.44E-11	3.10E-08	PPARG	upstream gene variant	ENSSSCG00000014224	SEMA6A
24	2	rs81363933	122,105	0.996	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81363986	122,538	0.004	1.45E-11	3.10E-08	PPARG	-	-	-
24	2	rs81225815	122,722	0.004	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81364080	123,299	0.996	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81364093	123,395	0.004	1.44E-11	3.10E-08	PPARG	intron variant	ENSSSCG00000028431	HSD17B4
24	2	rs81295472	123,998	0.996	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81364195	124,059	0.004	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81296107	124,163	0.996	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs80790446	124,860	0.004	1.44E-11	3.10E-08	PPARG	-	-	-
24	2	rs81474819	128,693	0.004	1.44E-11	3.10E-08	PPARG	-	-	-
25	2	rs81284541	119,937	0.004	3.30E-06	7.50E-03	PPARGC1A	intron variant	ENSSSCG00000014219	CDO1
25	2	rs81363852	120,690	0.004	3.30E-06	7.50E-03	PPARGC1A	upstream gene variant	ENSSSCG00000014224	SEMA6A
25	2	rs81363933	122,105	0.996	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81363986	122,538	0.004	3.16E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81225815	122,722	0.004	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81364080	123,299	0.996	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81364093	123,395	0.004	3.30E-06	7.50E-03	PPARGC1A	intron variant	ENSSSCG00000028431	HSD17B4
25	2	rs81295472	123,998	0.996	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81364195	124,059	0.004	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81296107	124,163	0.996	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs80790446	124,860	0.004	3.30E-06	7.50E-03	PPARGC1A	-	-	-
25	2	rs81474819	128,693	0.004	3.30E-06	7.50E-03	PPARGC1A	-	-	-
26	2	rs81284541	119,937	0.004	2.57E-09	5.49E-06	SCD	intron variant	ENSSSCG00000014219	CDO1
26	2	rs81363852	120,690	0.004	2.57E-09	5.49E-06	SCD	upstream gene variant	ENSSSCG00000014224	SEMA6A
26	2	rs81363933	122,105	0.996	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs81363986	122,538	0.004	2.50E-09	5.49E-06	SCD	-	-	-
26	2	rs81225815	122,722	0.004	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs81364080	123,299	0.996	2.57E-09	5.49E-06	SCD	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl Gene ID	Gene Symbol
26	2	rs81364093	123,395	0.004	2.57E-09	5.49E-06	SCD	intron variant	ENSSSCG00000028431	HSD17B4
26	2	rs81295472	123,998	0.996	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs81364195	124,059	0.004	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs81296107	124,163	0.996	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs80790446	124,860	0.004	2.57E-09	5.49E-06	SCD	-	-	-
26	2	rs81474819	128,693	0.004	2.57E-09	5.49E-06	SCD	-	-	-

BC1_DU:

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
1	1	rs81348429	13,163	0.885	1.18E-04	3.48E-02	ACSM5	-	-	-
1	1	rs81348541	13,630	0.139	1.12E-04	3.32E-02	ACSM5	upstream gene variant	ENSSSCG00000031720	-
1	1	rs81327383	14,456	0.07	2.02E-12	7.21E-08	ACSM5	intron variant	ENSSSCG00000025777	ESR1
2	1	rs80877479	246,699	0.787	9.52E-05	2.97E-02	ACSM5	intron variant	ENSSSCG00000005425	SLC44A1
2	1	rs81306994	250,592	0.086	2.60E-04	4.62E-02	ACSM5	-	-	-
2	1	rs80999779	258,899	0.77	1.11E-11	1.32E-07	ACSM5	-	-	-
2	1	rs80878629	258,930	0.23	1.11E-11	1.32E-07	ACSM5	-	-	-
3	2	rs81359986	8,713	0.086	2.60E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000028537	-
3	2	rs81360021	8,765	0.914	2.60E-04	4.62E-02	ACSM5	-	-	-
3	2	rs81368151	17,109	0.906	4.60E-07	3.21E-04	ACSM5	-	-	-
3	2	rs81368556	18,237	0.971	1.90E-04	4.22E-02	ACSM5	-	-	-
3	2	rs81335483	18,451	0.029	1.90E-04	4.22E-02	ACSM5	upstream gene variant	ENSSSCG00000013283	ACCSL
3	2	rs81257052	18,510	0.971	1.90E-04	4.22E-02	ACSM5	intron variant	ENSSSCG00000021739	HSD17B12
3	2	rs81238474	18,530	0.029	1.82E-04	4.13E-02	ACSM5	intron variant	ENSSSCG00000021739	HSD17B12
3	2	rs81355974	21,054	0.858	5.04E-05	1.74E-02	ACSM5	-	-	-
3	2	rs81304212	25,964	0.221	6.73E-05	2.22E-02	ACSM5	-	-	-
3	2	rs81271364	26,475	0.734	3.80E-05	1.36E-02	ACSM5	intron variant	ENSSSCG00000013301	ELF5
4	3	rs81324692	4,163	0.963	2.55E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000021285	RNF216
4	3	rs81269904	4,196	0.963	2.55E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000021285	RNF216
4	3	rs81242793	13,754	0.914	2.60E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000035984	-
4	3	rs80904058	16,240	0.932	8.34E-06	3.66E-03	ACSM5	-	-	-
4	3	rs81377182	19,351	0.23	2.93E-06	1.68E-03	ACSM5	intron variant	ENSSSCG00000007814	KIAA0556
4	3	rs81377282	19,382	0.23	2.93E-06	1.68E-03	ACSM5	intron variant	ENSSSCG00000007815	GTF3C1
4	3	rs81477684	22,920	0.045	6.39E-06	2.96E-03	ACSM5	-	-	-
4	3	rs81321464	23,166	0.139	1.51E-07	1.41E-04	ACSM5	-	-	-
4	3	rs81312070	23,335	0.914	2.87E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
4	3	rs81227560	23,359	0.086	2.87E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
4	3	rs341046181	23,433	0.861	1.51E-07	1.41E-04	ACSM5	-	-	-
4	3	rs81475137	23,446	0.111	6.08E-08	9.03E-05	ACSM5	-	-	-
4	3	rs81238947	23,519	0.139	1.51E-07	1.41E-04	ACSM5	-	-	-
4	3	rs81324887	23,568	0.889	6.08E-08	9.03E-05	ACSM5	downstream gene variant	ENSSSCG00000007838	OTOA
4	3	rs81340946	23,582	0.086	2.87E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000007838	OTOA
4	3	rs81313849	23,887	0.873	2.38E-09	7.88E-06	ACSM5	upstream gene variant	ENSSSCG00000020312	RF00026
4	3	rs81308074	23,896	0.14	1.56E-07	1.42E-04	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
4	3	rs323881880	23,900	0.139	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
4	3	rs81475068	23,968	0.913	2.69E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
4	3	rs81323675	24,582	0.086	2.87E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000034298	-
4	3	rs81278892	24,815	0.139	1.51E-07	1.41E-04	ACSM5	5 prime UTR variant	ENSSSCG00000007849	CRYM

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
4	3	rs81326798	24,892	0.861	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000031876	TMEM159
4	3	rs81336877	24,920	0.844	1.25E-06	8.09E-04	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
4	3	rs81288253	24,942	0.861	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
4	3	rs81313465	24,963	0.086	2.87E-09	7.88E-06	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
4	3	rs81315383	24,980	0.857	8.29E-05	2.64E-02	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
4	3	rs81379203	25,111	0.127	1.83E-07	1.59E-04	ACSM5	intron variant	ENSSSCG00000031798	DNAH3
4	3	rs81379199	25,155	0.861	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
4	3	rs81379197	25,173	0.14	1.88E-07	1.59E-04	ACSM5	intron variant	ENSSSCG00000007854	DCUN1D3
4	3	rs81315362	25,206	0.861	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000007855	REXO5
4	3	rs81288413	25,217	0.139	1.51E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000007855	REXO5
4	3	rs81335959	25,260	0.139	1.21E-07	1.41E-04	ACSM5	intron variant	ENSSSCG00000007857	ACSM3
4	3	rs81324695	25,404	0.843	1.59E-06	9.95E-04	ACSM5	intron variant	ENSSSCG00000026453	ACSM5
4	3	ACSM5.P	25,422	0.087	3.44E-09	8.76E-06	ACSM5	upstream gene variant	ENSSSCG00000026453	ACSM5
4	3	rs81326933	25,540	0.873	1.83E-07	1.59E-04	ACSM5	-	-	-
4	3	rs81347321	25,606	0.209	2.64E-05	9.81E-03	ACSM5	-	-	-
4	3	rs81278505	25,651	0.086	2.87E-09	7.88E-06	ACSM5	-	-	-
4	3	rs81240993	25,654	0.152	1.54E-06	9.78E-04	ACSM5	-	-	-
4	3	rs81379272	25,852	0.872	9.41E-08	1.24E-04	ACSM5	-	-	-
4	3	rs81239835	26,750	0.95	7.66E-05	2.50E-02	ACSM5	intron variant	ENSSSCG00000032531	SMG1
4	3	rs81238437	27,259	0.045	6.39E-06	2.96E-03	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
4	3	rs81379431	27,322	0.061	2.70E-05	9.83E-03	ACSM5	intron variant	ENSSSCG00000007872	XYLT1
4	3	rs81379567	27,603	0.905	5.28E-09	1.18E-05	ACSM5	-	-	-
4	3	rs81311765	31,176	0.119	1.68E-04	3.98E-02	ACSM5	-	-	-
4	3	rs81256954	32,830	0.033	2.32E-06	1.38E-03	ACSM5	-	-	-
4	3	rs81225882	32,847	0.033	2.32E-06	1.38E-03	ACSM5	-	-	-
5	4	rs80989635	9,484	0.25	2.62E-04	4.62E-02	ACSM5	-	-	-
5	4	rs80804375	9,505	0.75	2.62E-04	4.62E-02	ACSM5	-	-	-
5	4	rs81382229	9,697	0.738	2.10E-04	4.49E-02	ACSM5	intron variant	ENSSSCG00000031717	ADCY8
5	4	rs80855075	10,049	0.266	1.23E-04	3.53E-02	ACSM5	-	-	-
5	4	rs80853356	10,073	0.266	1.23E-04	3.53E-02	ACSM5	-	-	-
5	4	rs81001334	16,328	0.951	5.79E-07	3.96E-04	ACSM5	intron variant	ENSSSCG00000033690	DERL1
5	4	rs80979575	16,508	0.427	2.54E-04	4.62E-02	ACSM5	-	-	-
5	4	rs81294948	17,178	0.76	1.21E-04	3.52E-02	ACSM5	-	-	-
5	4	rs80812755	17,990	0.242	1.32E-04	3.74E-02	ACSM5	-	-	-
5	4	rs81337570	18,112	0.417	9.66E-05	2.97E-02	ACSM5	-	-	-
5	4	rs81323373	18,167	0.451	2.79E-04	4.87E-02	ACSM5	-	-	-
5	4	rs330213408	18,736	0.168	1.80E-04	4.13E-02	ACSM5	intron variant	ENSSSCG00000005997	COL14A1
6	6	rs81313468	2,754	0.885	2.38E-05	9.09E-03	ACSM5	-	-	-
6	6	rs81315523	4,428	0.951	1.60E-04	3.85E-02	ACSM5	downstream gene variant	ENSSSCG00000030446	KCNG4
6	6	rs81475968	4,528	0.049	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002682	MBTPS1
6	6	rs81306444	4,862	0.049	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81395086	5,321	0.951	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81393737	5,345	0.951	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81298997	5,370	0.95	1.49E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs336666922	5,522	0.951	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81294266	5,639	0.049	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81390670	5,662	0.049	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002684	CDH13
6	6	rs81239557	5,699	0.049	1.60E-04	3.85E-02	ACSM5	-	-	-
6	6	rs81394715	6,100	0.049	1.60E-04	3.85E-02	ACSM5	-	-	-
6	6	rs81393691	6,215	0.05	1.81E-04	4.13E-02	ACSM5	-	-	-
6	6	rs81393241	6,233	0.049	1.60E-04	3.85E-02	ACSM5	downstream gene variant	ENSSSCG00000020583	RF00100

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
6	6	rs81395073	6,467	0.95	1.64E-04	3.91E-02	ACSM5	intron variant	ENSSSCG00000002688	-
6	6	rs81395204	6,523	0.049	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002688	-
6	6	rs81395478	6,709	0.951	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002689	CMIP
6	6	rs81395631	6,971	0.951	1.60E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000002690	GAN
6	6	rs81339251	7,566	0.046	5.92E-05	1.99E-02	ACSM5	intron variant	ENSSSCG00000038128	CDYL2
7	7	rs80894452	42,263	0.094	2.20E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000030197	-
7	7	rs80967280	45,991	0.033	2.91E-04	4.99E-02	ACSM5	-	-	-
7	7	rs80797002	46,012	0.967	2.91E-04	4.99E-02	ACSM5	upstream gene variant	ENSSSCG00000001748	IL17A
7	7	rs80957448	46,257	0.033	2.91E-04	4.99E-02	ACSM5	intron variant	ENSSSCG00000002620	EFHC1
7	7	rs80936922	46,283	0.033	2.91E-04	4.99E-02	ACSM5	intron variant	ENSSSCG00000002620	EFHC1
8	11	rs81430408	22,645	0.902	8.78E-06	3.66E-03	ACSM5	intron variant	ENSSSCG00000038167	-
8	11	rs80884845	31,096	0.164	1.80E-04	4.13E-02	ACSM5	-	-	-
8	11	rs81430895	31,159	0.164	1.80E-04	4.13E-02	ACSM5	-	-	-
8	11	rs81224149	31,980	0.836	1.80E-04	4.13E-02	ACSM5	-	-	-
9	11	rs80825406	44,662	0.914	2.60E-04	4.62E-02	ACSM5	-	-	-
9	11	rs80952385	48,404	0.914	2.60E-04	4.62E-02	ACSM5	-	-	-
9	11	rs81307299	53,660	0.254	1.39E-04	3.85E-02	ACSM5	-	-	-
9	11	rs81320336	60,150	0.984	2.69E-05	9.83E-03	ACSM5	-	-	-
9	11	rs80882623	60,175	0.102	2.39E-04	4.62E-02	ACSM5	-	-	-
9	11	rs81431477	60,213	0.074	3.52E-08	5.69E-05	ACSM5	-	-	-
9	11	rs80829903	60,230	0.074	3.52E-08	5.69E-05	ACSM5	-	-	-
9	11	rs80933673	60,251	0.898	2.39E-04	4.62E-02	ACSM5	-	-	-
9	11	rs342798247	60,370	0.074	3.52E-08	5.69E-05	ACSM5	-	-	-
9	11	rs81326557	60,577	0.102	2.39E-04	4.62E-02	ACSM5	-	-	-
9	11	rs80876794	60,726	0.926	3.52E-08	5.69E-05	ACSM5	-	-	-
9	11	rs80907422	60,779	0.074	3.52E-08	5.69E-05	ACSM5	-	-	-
9	11	rs80914028	60,967	0.898	2.39E-04	4.62E-02	ACSM5	-	-	-
9	11	rs80972185	63,096	0.02	1.44E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000039364	-
9	11	rs80973460	63,137	0.893	3.81E-05	1.36E-02	ACSM5	intron variant	ENSSSCG00000039364	-
9	11	rs81344927	63,151	0.903	9.24E-07	6.10E-04	ACSM5	intron variant	ENSSSCG00000039364	-
9	11	rs81334585	63,153	0.898	2.40E-05	9.09E-03	ACSM5	intron variant	ENSSSCG00000039364	-
9	11	rs333523322	67,022	0.02	1.44E-04	3.85E-02	ACSM5	intron variant	ENSSSCG00000009509	IPO5
9	11	rs80908264	67,683	0.917	3.90E-07	2.82E-04	ACSM5	intron variant	ENSSSCG00000009513	SLC15A1
9	11	rs80812931	67,714	0.102	2.53E-04	4.62E-02	ACSM5	upstream gene variant	ENSSSCG00000009513	SLC15A1
9	11	rs81290322	69,242	0.053	5.19E-05	1.77E-02	ACSM5	intron variant	ENSSSCG00000009522	PCCA
9	11	rs80982881	73,572	0.98	1.44E-04	3.85E-02	ACSM5	-	-	-
9	11	rs81431996	73,794	0.979	1.92E-04	4.23E-02	ACSM5	-	-	-
9	11	rs80907085	73,817	0.02	1.44E-04	3.85E-02	ACSM5	-	-	-
9	11	rs80866439	74,148	0.02	1.44E-04	3.85E-02	ACSM5	-	-	-
9	11	rs344202101	74,642	0.186	2.03E-04	4.42E-02	ACSM5	-	-	-
9	11	rs80931687	76,803	0.211	1.91E-04	4.22E-02	ACSM5	-	-	-
10	12	rs81251734	59,035	0.058	3.63E-07	2.69E-04	ACSM5	upstream gene variant	ENSSSCG00000018031	ZNF287
10	12	rs81438317	59,267	0.057	2.58E-07	1.95E-04	ACSM5	intron variant	ENSSSCG00000037031	-
10	12	rs81344019	59,813	0.943	2.58E-07	1.95E-04	ACSM5	intron variant	ENSSSCG00000018045	ULK2
10	12	rs81235580	59,868	0.943	2.58E-07	1.95E-04	ACSM5	intron variant	ENSSSCG00000018045	ULK2
10	12	rs81289599	59,945	0.057	2.58E-07	1.95E-04	ACSM5	intron variant	ENSSSCG00000018041	ALDH3A2
11	13	rs80986028	175,321	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81441529	179,739	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81328764	186,269	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81478527	187,614	0.881	1.96E-06	1.21E-03	ACSM5	-	-	-
11	13	rs81340465	187,640	0.127	6.47E-05	2.15E-02	ACSM5	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
11	13	rs81327688	187,653	0.193	9.98E-05	2.99E-02	ACSM5	-	-	-
11	13	rs81267176	190,259	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81298331	190,267	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81442174	190,748	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81331679	190,797	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81331945	191,347	0.951	9.14E-06	3.66E-03	ACSM5	-	-	-
11	13	rs81272832	191,415	0.049	9.14E-06	3.66E-03	ACSM5	-	-	-
12	18	rs81321404	5,518	0.713	4.68E-06	2.35E-03	ACSM5	intron variant	ENSSSCG00000016432	PRKAG2
12	18	rs81312675	5,523	0.713	4.68E-06	2.35E-03	ACSM5	intron variant	ENSSSCG00000016432	PRKAG2
12	18	rs81332702	5,613	0.168	2.64E-04	4.64E-02	ACSM5	intron variant	ENSSSCG00000016432	PRKAG2
12	18	rs81237753	6,928	0.037	6.37E-06	2.96E-03	ACSM5	intron variant	ENSSSCG00000016462	CLCN1
12	18	rs81234760	8,006	0.963	6.37E-06	2.96E-03	ACSM5	intron variant	ENSSSCG00000031171	-
12	18	rs81256422	10,966	0.045	7.38E-07	4.96E-04	ACSM5	upstream gene variant	ENSSSCG00000018723	RF00019
12	18	rs81471951	15,290	0.963	6.77E-06	3.09E-03	ACSM5	intron variant	ENSSSCG00000016543	EXOC4
12	18	rs81467529	21,218	0.115	2.64E-05	9.81E-03	ACSM5	intron variant	ENSSSCG00000022865	GRM8
12	18	rs345497623	21,303	0.066	1.76E-11	1.57E-07	ACSM5	intron variant	ENSSSCG00000022865	GRM8
12	18	rs320703356	21,893	0.041	3.43E-06	1.91E-03	ACSM5	-	-	-
12	18	rs81248245	21,908	0.041	3.43E-06	1.91E-03	ACSM5	-	-	-
12	18	rs81467885	24,456	0.783	2.98E-04	5.06E-02	ACSM5	intron variant	ENSSSCG00000016611	CADPS2
12	18	rs81468173	26,156	0.279	2.06E-04	4.46E-02	ACSM5	-	-	-
12	18	rs81223574	31,258	0.045	1.45E-04	3.85E-02	ACSM5	-	-	-
12	18	rs81468822	34,873	0.963	4.23E-06	2.18E-03	ACSM5	-	-	-
12	18	rs81468830	34,909	0.037	4.23E-06	2.18E-03	ACSM5	-	-	-
12	18	rs81468835	34,927	0.963	4.23E-06	2.18E-03	ACSM5	-	-	-
12	18	rs81469168	38,145	0.929	9.85E-05	2.97E-02	ACSM5	intron variant	ENSSSCG00000016661	SEPT7
12	18	rs81469171	38,172	0.071	8.06E-05	2.59E-02	ACSM5	intron variant	ENSSSCG00000016661	SEPT7
12	18	rs81469206	38,606	0.037	4.23E-06	2.18E-03	ACSM5	-	-	-
12	18	rs81469204	38,630	0.963	4.23E-06	2.18E-03	ACSM5	-	-	-
12	18	rs81469437	41,452	0.889	4.89E-05	1.71E-02	ACSM5	downstream gene variant	ENSSSCG00000033535	PPP1R17
12	18	rs327181083	41,463	0.111	4.89E-05	1.71E-02	ACSM5	synonymous variant	ENSSSCG00000033535	PPP1R17
12	18	rs81246243	45,755	0.828	1.68E-04	3.98E-02	ACSM5	intron variant	ENSSSCG00000016708	SKAP2
12	18	rs323124537	46,209	0.225	2.41E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000036350	-
12	18	rs81479402	46,237	0.225	2.41E-04	4.62E-02	ACSM5	intron variant	ENSSSCG00000026894	NFE2L3
12	18	rs81470458	46,730	0.955	7.93E-09	1.66E-05	ACSM5	-	-	-
12	18	rs80902256	47,812	0.893	4.75E-09	1.13E-05	ACSM5	downstream gene variant	ENSSSCG00000038124	-
12	18	rs81297148	48,008	0.119	3.18E-10	2.27E-06	ACSM5	-	-	-
12	18	rs81470753	48,131	0.148	2.19E-07	1.82E-04	ACSM5	-	-	-
12	18	rs81245542	48,983	0.602	9.75E-05	2.97E-02	ACSM5	intron variant	ENSSSCG00000016725	TNS3
12	18	rs81471469	55,254	0.037	7.66E-08	1.09E-04	ACSM5	-	-	-
12	18	rs80811713	55,272	0.037	8.05E-08	1.10E-04	ACSM5	downstream gene variant	ENSSSCG00000040447	-
13	1	rs80843512	180,853	0.107	5.97E-08	5.32E-04	ACAA2	-	-	-
13	1	rs80846250	180,878	0.107	5.97E-08	5.32E-04	ACAA2	-	-	-
13	1	rs80926234	180,902	0.893	5.97E-08	5.32E-04	ACAA2	-	-	-
14	1	rs80910885	197,608	0.93	2.14E-07	6.36E-04	ACAA2	-	-	-
14	1	rs80799755	197,647	0.921	1.88E-07	6.36E-04	ACAA2	-	-	-
14	1	rs80835537	197,687	0.074	1.16E-06	2.95E-03	ACAA2	-	-	-
14	1	rs80810766	197,756	0.071	7.86E-08	5.60E-04	ACAA2	-	-	-
14	1	rs80886040	197,804	0.93	2.14E-07	6.36E-04	ACAA2	-	-	-
14	1	rs80997705	197,821	0.93	2.14E-07	6.36E-04	ACAA2	-	-	-
14	1	rs80904234	197,974	0.07	2.14E-07	6.36E-04	ACAA2	-	-	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
14	1	rs81350192	202,568	0.889	1.98E-06	4.42E-03	ACAA2	intron variant	ENSSSCG00000005166	MLLT3
14	1	rs81350204	202,619	0.111	1.98E-06	4.42E-03	ACAA2	-	-	-
14	1	rs80962874	203,367	0.869	1.16E-06	2.95E-03	ACAA2	-	-	-
14	1	rs80914376	203,632	0.09	3.97E-08	5.32E-04	ACAA2	upstream gene variant	ENSSSCG00000032205	-
15	1	rs80910885	197,608	0.93	2.09E-06	9.32E-03	CREG1	-	-	-
15	1	rs80799755	197,647	0.921	1.85E-08	6.59E-04	CREG1	-	-	-
15	1	rs80835537	197,687	0.074	1.19E-05	4.71E-02	CREG1	-	-	-
15	1	rs80810766	197,756	0.071	1.71E-06	9.32E-03	CREG1	-	-	-
15	1	rs80886040	197,804	0.93	2.09E-06	9.32E-03	CREG1	-	-	-
15	1	rs80997705	197,821	0.93	2.09E-06	9.32E-03	CREG1	-	-	-
15	1	rs80904234	197,974	0.07	2.09E-06	9.32E-03	CREG1	-	-	-
16	2	rs81315092	17,672	0.332	7.11E-06	3.20E-02	DGAT2	upstream gene variant	ENSSSCG00000013278	TSPAN18
16	2	rs81368350	17,683	0.668	7.11E-06	3.20E-02	DGAT2	-	-	-
16	2	rs81270847	25,107	0.557	1.18E-05	4.27E-02	DGAT2	intron variant	ENSSSCG00000013294	LDLRAD3
16	2	rs81240674	25,446	0.668	2.27E-06	2.02E-02	DGAT2	-	-	-
16	2	rs81347362	25,467	0.164	8.34E-07	1.09E-02	DGAT2	-	-	-
16	2	rs81287004	26,204	0.554	7.19E-06	3.20E-02	DGAT2	-	-	-
16	2	rs81254068	26,400	0.553	1.20E-05	4.27E-02	DGAT2	-	-	-
17	9	rs81408950	34,248	0.996	7.04E-06	2.09E-02	ETS1	-	-	-
17	9	rs81408951	34,269	0.004	7.04E-06	2.09E-02	ETS1	-	-	-
17	9	rs81409949	42,607	0.996	7.04E-06	2.09E-02	ETS1	-	-	-
17	9	rs81409945	42,741	0.004	7.04E-06	2.09E-02	ETS1	intron variant	ENSSSCG00000031831	CADM1
17	9	rs81409997	42,796	0.996	7.04E-06	2.09E-02	ETS1	intron variant	ENSSSCG00000031831	CADM1
17	9	rs81410023	42,977	0.004	7.04E-06	2.09E-02	ETS1	intron variant	ENSSSCG00000031831	CADM1
17	9	rs81410136	43,420	0.996	7.04E-06	2.09E-02	ETS1	-	-	-
17	9	rs81410215	43,705	0.996	7.04E-06	2.09E-02	ETS1	-	-	-
17	9	rs81410468	44,486	0.996	7.04E-06	2.09E-02	ETS1	-	-	-
18	2	rs81306755	145	0.388	2.06E-16	1.05E-12	IGF2	intron variant	ENSSSCG00000014565	-
18	2	rs81328276	236	0.32	1.21E-06	7.19E-04	IGF2	intron variant	ENSSSCG00000024569	ANO9
18	2	rs81317307	310	0.385	8.46E-17	5.75E-13	IGF2	intron variant	ENSSSCG00000027045	LRRC56
18	2	rs81339115	422	0.318	6.60E-07	4.61E-04	IGF2	intron variant	ENSSSCG00000012850	DEAF1
18	2	rs81341763	677	0.318	2.06E-07	1.63E-04	IGF2	-	-	-
18	2	IGF2	1,483	0.275	1.31E-18	4.66E-14	IGF2	-	-	-
18	2	rs81291529	2,636	0.393	1.66E-17	1.98E-13	IGF2	intron variant	ENSSSCG00000033043	SHANK2
18	2	rs81336288	3,062	0.824	9.45E-06	4.16E-03	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81237841	3,681	0.463	1.95E-10	3.47E-07	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81322199	3,689	0.316	3.10E-05	1.19E-02	IGF2	intron variant	ENSSSCG00000031191	-
18	2	rs81357266	3,985	0.376	6.56E-17	5.75E-13	IGF2	-	-	-
18	2	rs81361514	4,341	0.742	2.80E-14	7.13E-11	IGF2	intron variant	ENSSSCG00000032760	-
18	2	rs81363333	4,378	0.68	2.41E-15	6.60E-12	IGF2	-	-	-
18	2	rs81364067	4,412	0.701	3.40E-14	7.58E-11	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
18	2	rs81364734	4,444	0.701	3.40E-14	7.58E-11	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
18	2	rs81337384	4,531	0.795	9.09E-12	1.80E-08	IGF2	downstream gene variant	ENSSSCG00000012885	-
18	2	rs81237341	4,671	0.107	2.43E-05	9.61E-03	IGF2	-	-	-
18	2	rs81368356	4,795	0.902	9.70E-07	6.52E-04	IGF2	intron variant	ENSSSCG00000012889	CHKA
18	2	rs81367772	4,892	0.652	1.45E-04	4.47E-02	IGF2	upstream gene variant	ENSSSCG00000028501	-
18	2	rs81368610	4,930	0.098	9.70E-07	6.52E-04	IGF2	downstream gene variant	ENSSSCG00000026349	ALDH3B2
18	2	rs81356358	5,289	0.545	1.67E-11	3.13E-08	IGF2	intron variant	ENSSSCG00000029637	KDM2A
18	2	rs81356578	5,840	0.303	1.32E-07	1.12E-04	IGF2	downstream gene variant	ENSSSCG00000012933	CCDC87

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
18	2	rs81357066	6,129	0.426	1.16E-04	3.73E-02	IGF2	upstream gene variant	ENSSSCG00000012955	KLC2
18	2	rs81357081	6,150	0.087	2.81E-06	1.47E-03	IGF2	intron variant	ENSSSCG00000012956	PACS1
18	2	rs81357172	6,263	0.07	4.23E-05	1.57E-02	IGF2	-	-	-
18	2	rs81271991	6,640	0.537	4.17E-09	5.94E-06	IGF2	upstream gene variant	ENSSSCG00000012983	PCNX3
18	2	rs81358530	6,962	0.799	2.55E-07	1.97E-04	IGF2	intron variant	ENSSSCG00000012999	CAPN1
18	2	rs81343625	7,199	0.627	4.57E-06	2.14E-03	IGF2	-	-	-
18	2	rs81290024	7,244	0.398	7.07E-05	2.40E-02	IGF2	upstream gene variant	ENSSSCG00000013015	GPHA2
18	2	rs81246704	7,324	0.373	4.57E-06	2.14E-03	IGF2	upstream gene variant	ENSSSCG00000013018	CDC42BPG
18	2	rs81343787	7,363	0.322	4.22E-18	7.52E-14	IGF2	intron variant	ENSSSCG00000013019	MEN1
18	2	rs81333747	7,394	0.627	4.57E-06	2.14E-03	IGF2	intron variant	ENSSSCG00000013021	SF1
18	2	rs81323771	7,449	0.475	3.07E-07	2.33E-04	IGF2	-	-	-
18	2	rs81359237	7,705	0.414	2.44E-08	2.72E-05	IGF2	-	-	-
18	2	rs81212188	7,829	0.824	5.25E-05	1.85E-02	IGF2	synonymous variant	ENSSSCG00000013032	GPR137
18	2	rs81271004	8,368	0.627	7.13E-05	2.40E-02	IGF2	intron variant	ENSSSCG00000040120	-
18	2	rs81285409	8,393	0.533	1.13E-06	7.19E-04	IGF2	upstream gene variant	ENSSSCG00000040120	-
18	2	rs81359894	8,540	0.664	1.18E-05	4.94E-03	IGF2	-	-	-
18	2	rs81360111	8,647	0.5	1.09E-06	7.19E-04	IGF2	intron variant	ENSSSCG00000028537	-
18	2	rs81359966	8,687	0.574	4.80E-05	1.71E-02	IGF2	intron variant	ENSSSCG00000028537	-
18	2	rs81474907	8,795	0.766	7.10E-05	2.40E-02	IGF2	intron variant	ENSSSCG00000029516	SLC22A8
18	2	rs81314686	9,226	0.266	4.76E-05	1.71E-02	IGF2	intron variant	ENSSSCG00000036669	-
18	2	rs325325237	9,419	0.412	1.63E-07	1.35E-04	IGF2	intron variant	ENSSSCG00000013066	INCENP
18	2	rs81474400	9,435	0.439	5.23E-09	6.91E-06	IGF2	intron variant	ENSSSCG00000013066	INCENP
18	2	rs81308303	9,495	0.59	4.58E-09	6.27E-06	IGF2	-	-	-
18	2	rs81360403	9,588	0.413	1.03E-08	1.18E-05	IGF2	intron variant	ENSSSCG00000013074	RAB3IL1
18	2	rs331754883	9,860	0.258	3.73E-08	3.49E-05	IGF2	intron variant	ENSSSCG00000013079	DAGLA
18	2	rs81360839	10,031	0.455	5.13E-06	2.37E-03	IGF2	-	-	-
18	2	rs81361056	10,099	0.332	1.13E-09	1.68E-06	IGF2	intron variant	ENSSSCG00000013083	CPSF7
18	2	rs81361093	10,135	0.287	8.70E-08	7.95E-05	IGF2	3 prime UTR variant	ENSSSCG00000037843	TMEM138
18	2	rs81273412	10,230	0.852	2.48E-10	4.20E-07	IGF2	intron variant	ENSSSCG00000013093	VWCE
18	2	rs81361441	10,521	0.295	4.62E-07	3.36E-04	IGF2	intron variant	ENSSSCG00000013097	-
18	2	rs81361464	10,543	0.619	8.02E-09	9.86E-06	IGF2	intron variant	ENSSSCG00000029938	-
18	2	rs81361375	10,584	0.619	8.02E-09	9.86E-06	IGF2	upstream gene variant	ENSSSCG00000029938	-
18	2	rs81362098	10,626	0.43	1.98E-05	8.03E-03	IGF2	intron variant	ENSSSCG00000028762	VPS37C
18	2	rs81362233	10,687	0.447	3.89E-06	1.93E-03	IGF2	-	-	-
18	2	rs81362189	10,739	0.361	4.25E-07	3.16E-04	IGF2	3 prime UTR variant	ENSSSCG00000013111	CD6
18	2	rs81362046	10,771	0.582	1.20E-05	4.96E-03	IGF2	intron variant	ENSSSCG00000013111	CD6
18	2	rs81361857	10,828	0.375	5.47E-06	2.50E-03	IGF2	downstream gene variant	ENSSSCG00000013110	TMEM109
18	2	rs81361790	10,834	0.406	1.23E-06	7.19E-04	IGF2	intron variant	ENSSSCG00000013110	TMEM109
18	2	rs80984785	10,869	0.631	8.53E-06	3.80E-03	IGF2	-	-	-
18	2	rs81246105	10,892	0.406	1.23E-06	7.19E-04	IGF2	intron variant	ENSSSCG00000013106	PTGDR2
18	2	rs81361933	10,910	0.409	2.45E-06	1.30E-03	IGF2	upstream gene variant	ENSSSCG00000013105	CCDC86
18	2	rs81362017	10,951	0.382	6.13E-07	4.37E-04	IGF2	intron variant	ENSSSCG00000037987	-
18	2	rs81361972	10,965	0.594	1.23E-06	7.19E-04	IGF2	intron variant	ENSSSCG00000034847	-
18	2	rs330591156	11,057	0.352	1.35E-06	7.74E-04	IGF2	downstream gene variant	ENSSSCG00000031637	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
18	2	rs81361668	11,078	0.594	1.23E-06	7.19E-04	IGF2	-	-	-
18	2	rs81361684	11,120	0.594	1.23E-06	7.19E-04	IGF2	-	-	-
18	2	rs81361680	11,132	0.358	1.44E-05	5.91E-03	IGF2	-	-	-
18	2	rs81338206	11,253	0.156	2.72E-08	2.74E-05	IGF2	intron variant	ENSSSCG00000026796	MS4A5
18	2	rs81324813	11,268	0.632	1.02E-08	1.18E-05	IGF2	downstream gene variant	ENSSSCG00000022114	-
18	2	rs81361507	11,440	0.721	9.84E-16	2.92E-12	IGF2	-	-	-
18	2	rs81278022	11,580	0.723	9.69E-17	5.75E-13	IGF2	-	-	-
18	2	rs81336616	11,627	0.721	9.84E-16	2.92E-12	IGF2	-	-	-
18	2	rs81240151	11,668	0.451	4.51E-06	2.14E-03	IGF2	intron variant	ENSSSCG00000013117	CBLIF
18	2	rs81474697	11,683	0.279	9.84E-16	2.92E-12	IGF2	intron variant	ENSSSCG00000013119	STX3
18	2	rs81253085	11,693	0.279	9.84E-16	2.92E-12	IGF2	intron variant	ENSSSCG00000013118	MRPL16
18	2	rs81324228	11,711	0.193	1.13E-07	9.82E-05	IGF2	intron variant	ENSSSCG00000013119	STX3
18	2	rs81322356	11,716	0.279	9.84E-16	2.92E-12	IGF2	intron variant	ENSSSCG00000013119	STX3
18	2	rs81257178	11,836	0.816	4.15E-05	1.56E-02	IGF2	intron variant	ENSSSCG00000013124	PATL1
18	2	rs81238148	11,975	0.678	9.22E-08	8.21E-05	IGF2	upstream gene variant	ENSSSCG00000031679	-
18	2	rs81345516	12,209	0.102	4.30E-05	1.58E-02	IGF2	intron variant	ENSSSCG00000013145	DTX4
18	2	rs81474708	12,225	0.318	8.58E-05	2.86E-02	IGF2	intron variant	ENSSSCG00000013145	DTX4
18	2	rs81331133	12,718	0.619	1.80E-06	9.82E-04	IGF2	upstream gene variant	ENSSSCG00000039026	-
18	2	rs81305360	12,725	0.381	1.80E-06	9.82E-04	IGF2	downstream gene variant	ENSSSCG00000039026	-
18	2	rs81322752	12,817	0.553	2.71E-05	1.05E-02	IGF2	-	-	-
18	2	rs81362768	12,935	0.422	4.59E-05	1.67E-02	IGF2	-	-	-
18	2	rs81262060	12,997	0.381	1.80E-06	9.82E-04	IGF2	upstream gene variant	ENSSSCG00000013170	-
18	2	rs81362978	13,105	0.656	2.84E-08	2.74E-05	IGF2	-	-	-
18	2	rs332366314	13,156	0.303	4.70E-13	9.84E-10	IGF2	3 prime UTR variant	ENSSSCG00000013174	CTNND1
18	2	rs318322737	13,167	0.344	2.84E-08	2.74E-05	IGF2	synonymous variant	ENSSSCG00000013174	CTNND1
18	2	rs81363153	13,192	0.344	2.84E-08	2.74E-05	IGF2	intron variant	ENSSSCG00000013174	CTNND1
18	2	rs81363209	13,218	0.656	2.84E-08	2.74E-05	IGF2	-	-	-
18	2	rs81363250	13,230	0.661	2.06E-07	1.63E-04	IGF2	3 prime UTR variant	ENSSSCG00000013176	-
18	2	rs81363413	13,307	0.624	7.66E-10	1.19E-06	IGF2	-	-	-
18	2	rs332503014	13,334	0.31	1.04E-05	4.48E-03	IGF2	intron variant	ENSSSCG00000013181	SERPING1
18	2	rs81305603	13,392	0.602	1.33E-04	4.12E-02	IGF2	upstream gene variant	ENSSSCG00000039659	TIMM10
18	2	rs81340329	13,620	0.512	3.68E-10	5.95E-07	IGF2	upstream gene variant	ENSSSCG00000029468	P2RX3
18	2	rs81363706	13,970	0.738	1.03E-05	4.46E-03	IGF2	intron variant	ENSSSCG00000029871	-
18	2	rs81363838	14,000	0.205	6.90E-05	2.40E-02	IGF2	upstream gene variant	ENSSSCG00000013193	-
18	2	rs81365207	15,189	0.807	1.01E-04	3.30E-02	IGF2	downstream gene variant	ENSSSCG00000013238	SLC39A13
18	2	rs81367209	16,404	0.607	2.55E-05	1.00E-02	IGF2	intron variant	ENSSSCG00000023974	PHF21A
18	2	rs81295533	16,426	0.668	2.04E-05	8.17E-03	IGF2	intron variant	ENSSSCG00000023974	PHF21A
18	2	rs81336030	16,902	0.652	1.82E-06	9.82E-04	IGF2	-	-	-
18	2	rs81319870	16,905	0.237	9.57E-05	3.16E-02	IGF2	-	-	-
18	2	rs81474772	17,633	0.426	1.31E-04	4.12E-02	IGF2	intron variant	ENSSSCG00000013278	TSPAN18
18	2	rs81474454	17,640	0.574	1.31E-04	4.12E-02	IGF2	intron variant	ENSSSCG00000013278	TSPAN18
19	4	rs80929170	107,415	0.352	6.25E-06	2.82E-03	IGF2	intron variant	ENSSSCG00000034581	LRIG2
19	4	rs81312576	110,464	0.347	3.71E-06	1.86E-03	IGF2	intron variant	ENSSSCG00000037808	-

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
19	4	rs81380274	110,644	0.665	2.86E-06	1.48E-03	IGF2	downstream gene variant	ENSSSCG00000006829	SYPL2
19	4	rs81380278	110,658	0.668	3.62E-06	1.84E-03	IGF2	intron variant	ENSSSCG00000006829	SYPL2
19	4	rs81380295	110,708	0.676	1.13E-05	4.78E-03	IGF2	intron variant	ENSSSCG00000006830	-
19	4	rs341806732	110,758	0.594	1.32E-04	4.12E-02	IGF2	synonymous variant	ENSSSCG00000006831	SORT1
20	4	rs80876714	35,029	0.5	4.00E-05	5.09E-02	LPIN1	intron variant	ENSSSCG00000006059	NCALD
20	4	rs81001300	36,454	0.156	2.68E-05	4.82E-02	LPIN1	-	-	-
20	4	rs80856006	40,165	0.336	1.28E-05	4.14E-02	LPIN1	-	-	-
20	4	rs81223915	48,234	0.155	3.15E-05	4.82E-02	LPIN1	-	-	-
20	4	rs80842911	48,336	0.164	3.57E-05	4.82E-02	LPIN1	-	-	-
21	4	rs80927846	67,460	0.893	3.36E-05	4.82E-02	LPIN1	upstream gene variant	ENSSSCG00000006201	ARFGEF1
21	4	rs80971648	67,504	0.107	3.45E-05	4.82E-02	LPIN1	intron variant	ENSSSCG00000006201	ARFGEF1
21	4	rs81382125	71,023	0.758	2.34E-05	4.82E-02	LPIN1	downstream gene variant	ENSSSCG00000037637	GGH
21	4	rs80963244	76,419	0.623	3.43E-05	4.82E-02	LPIN1	-	-	-
21	4	rs80936556	76,574	0.652	2.22E-05	4.82E-02	LPIN1	-	-	-
21	4	rs80856841	76,599	0.652	2.22E-05	4.82E-02	LPIN1	-	-	-
21	4	rs80933685	84,642	0.583	3.65E-05	4.82E-02	LPIN1	intron variant	ENSSSCG00000006321	FAM78B
22	7	rs81266661	108,190	0.701	1.16E-06	7.77E-03	LPIN1	-	-	-
22	7	rs80812481	111,423	0.238	2.64E-05	4.82E-02	LPIN1	intron variant	ENSSSCG00000002429	FOXN3
22	7	rs80793518	111,428	0.24	2.90E-05	4.82E-02	LPIN1	intron variant	ENSSSCG00000002429	FOXN3
23	15	rs80798447	103,751	0.205	1.20E-06	7.77E-03	LPIN1	-	-	-
23	15	rs81454150	103,867	0.205	1.20E-06	7.77E-03	LPIN1	-	-	-
23	15	rs81454155	103,901	0.208	1.53E-06	7.77E-03	LPIN1	intron variant	ENSSSCG00000016090	SPATS2L
23	15	rs81454170	104,022	0.795	1.20E-06	7.77E-03	LPIN1	intron variant	ENSSSCG00000016090	SPATS2L
23	15	rs81454180	104,067	0.795	1.20E-06	7.77E-03	LPIN1	downstream gene variant	ENSSSCG00000016091	-
23	15	rs81254527	104,093	0.793	1.40E-06	7.77E-03	LPIN1	intron variant	ENSSSCG00000016091	-
24	1	rs80910885	197,608	0.93	6.69E-07	3.41E-03	NCOA1	-	-	-
24	1	rs80835537	197,687	0.074	9.62E-07	4.28E-03	NCOA1	-	-	-
24	1	rs80810766	197,756	0.071	4.91E-07	3.41E-03	NCOA1	-	-	-
24	1	rs80886040	197,804	0.93	6.69E-07	3.41E-03	NCOA1	-	-	-
24	1	rs80997705	197,821	0.93	6.69E-07	3.41E-03	NCOA1	-	-	-
24	1	rs80904234	197,974	0.07	6.69E-07	3.41E-03	NCOA1	-	-	-
24	1	rs80914376	203,632	0.09	2.21E-06	8.74E-03	NCOA1	upstream gene variant	ENSSSCG00000032205	-
25	1	rs80910885	197,608	0.93	3.34E-07	1.49E-03	NCOA6	-	-	-
25	1	rs80799755	197,647	0.921	2.07E-06	7.37E-03	NCOA6	-	-	-
25	1	rs80835537	197,687	0.074	1.74E-07	1.49E-03	NCOA6	-	-	-
25	1	rs80810766	197,756	0.071	1.78E-07	1.49E-03	NCOA6	-	-	-
25	1	rs80886040	197,804	0.93	3.34E-07	1.49E-03	NCOA6	-	-	-
25	1	rs80997705	197,821	0.93	3.34E-07	1.49E-03	NCOA6	-	-	-
25	1	rs80904234	197,974	0.07	3.34E-07	1.49E-03	NCOA6	-	-	-
25	1	rs81350192	202,568	0.889	1.08E-05	3.06E-02	NCOA6	intron variant	ENSSSCG00000005166	MLLT3
25	1	rs81350204	202,619	0.111	1.08E-05	3.06E-02	NCOA6	-	-	-
25	1	rs80914376	203,632	0.09	1.12E-05	3.06E-02	NCOA6	upstream gene variant	ENSSSCG00000032205	-
26	1	rs81322865	159,785	0.471	7.59E-06	1.18E-02	PDHX	non coding transcript variant	ENSSSCG00000034988	-
26	1	rs323537067	159,835	0.529	7.59E-06	1.18E-02	PDHX	intron variant	ENSSSCG00000004903	CDH20
26	1	rs80975790	159,873	0.471	7.59E-06	1.18E-02	PDHX	intron variant	ENSSSCG00000004903	CDH20

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
26	1	rs80979649	167,660	0.93	2.00E-07	5.49E-04	PDHX	non coding transcript variant	ENSSSCG00000037389	-
26	1	rs81349537	167,898	0.111	1.00E-05	1.49E-02	PDHX	-	-	-
26	1	rs81326119	167,989	0.07	2.00E-07	5.49E-04	PDHX	-	-	-
27	1	rs80992169	180,585	0.295	1.27E-05	1.66E-02	PDHX	-	-	-
27	1	rs80843512	180,853	0.107	1.59E-07	5.49E-04	PDHX	-	-	-
27	1	rs80846250	180,878	0.107	1.59E-07	5.49E-04	PDHX	-	-	-
27	1	rs80926234	180,902	0.893	1.59E-07	5.49E-04	PDHX	-	-	-
27	1	rs81349785	182,905	0.164	7.56E-06	1.18E-02	PDHX	-	-	-
27	1	rs80881542	185,683	0.123	1.29E-06	2.71E-03	PDHX	-	-	-
27	1	rs81349921	190,103	0.115	1.30E-05	1.66E-02	PDHX	-	-	-
27	1	rs80910885	197,608	0.93	2.00E-07	5.49E-04	PDHX	-	-	-
27	1	rs80799755	197,647	0.921	2.67E-05	3.17E-02	PDHX	-	-	-
27	1	rs80835537	197,687	0.074	6.73E-07	1.50E-03	PDHX	-	-	-
27	1	rs80810766	197,756	0.071	1.03E-07	5.49E-04	PDHX	-	-	-
27	1	rs80886040	197,804	0.93	2.00E-07	5.49E-04	PDHX	-	-	-
27	1	rs80997705	197,821	0.93	2.00E-07	5.49E-04	PDHX	-	-	-
27	1	rs80904234	197,974	0.07	2.00E-07	5.49E-04	PDHX	-	-	-
27	1	rs81350192	202,568	0.889	1.44E-07	5.49E-04	PDHX	intron variant	ENSSSCG00000005166	MLLT3
27	1	rs81350204	202,619	0.111	1.44E-07	5.49E-04	PDHX	-	-	-
27	1	rs81350198	202,703	0.872	1.77E-05	2.17E-02	PDHX	-	-	-
27	1	rs337871351	202,846	0.115	1.30E-05	1.66E-02	PDHX	-	-	-
27	1	rs80933033	202,874	0.885	1.30E-05	1.66E-02	PDHX	-	-	-
27	1	rs80962874	203,367	0.869	2.91E-07	7.41E-04	PDHX	-	-	-
27	1	rs80914376	203,632	0.09	1.91E-08	5.49E-04	PDHX	upstream gene variant	ENSSSCG00000032205	-
28	2	rs81358527	43,500	0.201	1.42E-05	2.49E-02	PPARA	-	-	-
28	2	rs81358560	43,603	0.205	1.15E-05	2.42E-02	PPARA	-	-	-
28	2	rs81343459	43,686	0.205	1.15E-05	2.42E-02	PPARA	-	-	-
29	15	rs80798447	103,751	0.205	5.03E-07	2.24E-03	PPARA	-	-	-
29	15	rs81454150	103,867	0.205	5.03E-07	2.24E-03	PPARA	-	-	-
29	15	rs81454155	103,901	0.208	4.43E-07	2.24E-03	PPARA	intron variant	ENSSSCG00000016090	SPATS2L
29	15	rs81454170	104,022	0.795	5.03E-07	2.24E-03	PPARA	intron variant	ENSSSCG00000016090	SPATS2L
29	15	rs81454180	104,067	0.795	5.03E-07	2.24E-03	PPARA	downstream gene variant	ENSSSCG00000016091	-
29	15	rs81254527	104,093	0.793	4.93E-07	2.24E-03	PPARA	intron variant	ENSSSCG00000016091	-
30	17	rs81465871	29,976	0.779	4.32E-07	2.24E-03	PPARA	-	-	-
30	17	rs80870918	32,140	0.779	1.43E-05	2.49E-02	PPARA	intron variant	ENSSSCG00000007153	ATRN
30	17	rs80898068	32,153	0.221	1.43E-05	2.49E-02	PPARA	intron variant	ENSSSCG00000007153	ATRN
30	17	rs80949545	32,183	0.754	1.92E-06	6.21E-03	PPARA	intron variant	ENSSSCG00000007153	ATRN
30	17	rs80938594	34,461	0.75	1.03E-05	2.42E-02	PPARA	intron variant	ENSSSCG00000007199	SLC52A3
31	1	rs80910885	197,608	0.93	5.90E-06	2.10E-02	PRKAA1	-	-	-
31	1	rs80799755	197,647	0.921	7.94E-07	2.10E-02	PRKAA1	-	-	-
31	1	rs80835537	197,687	0.074	7.73E-06	2.50E-02	PRKAA1	-	-	-
31	1	rs80810766	197,756	0.071	4.22E-06	2.10E-02	PRKAA1	-	-	-
31	1	rs80886040	197,804	0.93	5.90E-06	2.10E-02	PRKAA1	-	-	-
31	1	rs80997705	197,821	0.93	5.90E-06	2.10E-02	PRKAA1	-	-	-
31	1	rs80904234	197,974	0.07	5.90E-06	2.10E-02	PRKAA1	-	-	-
32	4	rs80856316	2,524	0.533	1.31E-06	1.16E-02	PXMP3	downstream gene variant	ENSSSCG00000005930	SLC45A4
32	4	rs80889654	2,620	0.533	1.31E-06	1.16E-02	PXMP3	-	-	-
32	4	rs80893032	2,648	0.533	1.31E-06	1.16E-02	PXMP3	-	-	-
32	4	rs80801576	2,905	0.533	1.31E-06	1.16E-02	PXMP3	-	-	-

BC1_PI:

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
1	1	rs81233254	271,398	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000005715	PRRC2B
1	1	rs81285030	271,664	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
1	1	rs81353054	271,701	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
1	1	rs80881914	271,830	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000005720	MED27
1	1	rs80994324	271,852	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000005720	MED27
1	1	rs80893612	271,889	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000005720	MED27
1	1	rs80923749	271,892	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000005720	MED27
1	1	rs341500950	271,933	0.004	1.64E-05	8.84E-03	ACSM5	upstream gene variant	ENSSSCG00000005720	MED27
2	3	rs81293818	22,810	0.894	8.41E-05	4.05E-02	ACSM5	intron variant	ENSSSCG00000010799	COG7
2	3	rs81312070	23,335	0.847	1.80E-06	8.84E-03	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
2	3	rs81227560	23,359	0.153	1.80E-06	8.84E-03	ACSM5	intron variant	ENSSSCG00000036475	HS3ST2
2	3	rs81475068	23,968	0.898	7.32E-09	8.82E-05	ACSM5	intron variant	ENSSSCG00000007839	EEF2K
2	3	ACSM5.P	25,422	0.102	7.32E-09	8.82E-05	ACSM5	upstream gene variant	ENSSSCG00000026453	ACSM5
2	3	rs81347321	25,606	0.318	1.94E-05	9.99E-03	ACSM5	-	-	-
2	3	rs81278505	25,651	0.102	7.32E-09	8.82E-05	ACSM5	-	-	-
3	3	rs81371705	64,193	0.157	8.57E-05	4.07E-02	ACSM5	intron variant	ENSSSCG00000008254	-
3	3	rs81371763	64,831	0.237	3.31E-05	1.66E-02	ACSM5	intron variant	ENSSSCG00000037620	-
3	3	rs81371921	67,378	0.216	4.13E-05	2.02E-02	ACSM5	-	-	-
3	3	rs81313500	67,608	0.216	4.13E-05	2.02E-02	ACSM5	intron variant	ENSSSCG00000021007	MRPL19
4	8	rs81402998	103,957	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81403010	104,091	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81403022	104,775	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009111	SYNPO2
4	8	rs81403067	105,054	0.996	1.64E-05	8.84E-03	ACSM5	upstream gene variant	ENSSSCG00000009113	METTL14
4	8	rs81255350	105,251	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009115	NDST3
4	8	rs81403203	105,770	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81263179	106,616	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81306885	108,244	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000031904	UGT8
4	8	rs81477042	108,307	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000031904	UGT8
4	8	rs81403300	108,767	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009122	ARSJ
4	8	rs81403305	108,799	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009122	ARSJ
4	8	rs81403308	108,826	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81403315	108,875	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81403328	108,957	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009123	CAMK2D
4	8	rs81403331	108,978	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009123	CAMK2D
4	8	rs81338904	109,230	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81339074	109,317	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009125	ANK2
4	8	rs81273257	109,319	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009125	ANK2
4	8	rs81315019	109,322	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009125	ANK2
4	8	rs334851184	109,331	0.996	1.64E-05	8.84E-03	ACSM5	synonymous variant	ENSSSCG00000009125	ANK2
4	8	rs81403348	109,399	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009125	ANK2
4	8	rs81403355	109,446	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009125	ANK2
4	8	rs81403368	109,576	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81477002	113,809	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81332214	114,403	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81306425	114,474	0.996	1.66E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81292625	114,483	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
4	8	rs81301569	116,475	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009157	TET2

Inter-val	Chr	SNP	Position (Mb)	MAF	p-value	FDR	Associated Gene	Consequence	Ensembl GeneID	Gene Symbol
4	8	rs81329186	116,677	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
5	12	rs81271493	27,551	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000017593	UTP18
5	12	rs81433435	31,562	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
5	12	rs81293225	31,648	0.996	1.64E-05	8.84E-03	ACSM5	downstream gene variant	ENSSSCG00000017604	HLF
6	14	rs80785686	14,101	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009693	XKR6
6	14	rs81000143	14,339	0.996	1.64E-05	8.84E-03	ACSM5	-	-	-
6	14	rs81330598	14,743	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000027568	BLK
6	14	rs81325322	14,921	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000022383	GATA4
6	14	rs80959233	16,765	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000009705	GALNT7
6	14	rs332515193	17,897	0.004	1.64E-05	8.84E-03	ACSM5	-	-	-
7	16	rs81460917	62,156	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000033558	GABRB2
7	16	rs81460922	62,183	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000033558	GABRB2
7	16	rs81460983	63,245	0.004	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000017036	CCNJL
7	16	rs81462197	71,272	0.996	1.64E-05	8.84E-03	ACSM5	intron variant	ENSSSCG00000035676	G3BP1
8	7	rs80913379	111,283	0.366	2.48E-05	3.59E-02	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
8	7	rs80807511	111,328	0.639	1.71E-05	2.57E-02	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
8	7	rs326024106	111,384	0.412	3.05E-05	4.24E-02	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
8	7	rs80812481	111,423	0.403	2.01E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
8	7	rs80793518	111,428	0.403	2.01E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002429	FOXN3
8	7	rs80938538	111,492	0.605	2.66E-07	5.12E-03	ACSS2	-	-	-
8	7	rs80871598	111,558	0.496	2.27E-06	5.12E-03	ACSS2	-	-	-
8	7	rs81396214	111,780	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002431	TDP1
8	7	rs81396221	111,804	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002431	TDP1
8	7	rs81396243	111,852	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002431	TDP1
8	7	rs81396246	111,869	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002432	KCNK13
8	7	rs81396256	111,911	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002432	KCNK13
8	7	rs81396277	111,934	0.504	2.27E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002432	KCNK13
8	7	rs80970460	111,999	0.5	1.59E-06	5.12E-03	ACSS2	-	-	-
8	7	rs80808783	112,013	0.504	2.27E-06	5.12E-03	ACSS2	-	-	-
8	7	rs80839580	112,044	0.496	2.27E-06	5.12E-03	ACSS2	3 prime UTR variant	ENSSSCG00000002433	PSMC1
8	7	rs80887503	112,083	0.496	1.12E-06	5.12E-03	ACSS2	intron variant	ENSSSCG00000002434	NRDE2
8	7	rs81001496	112,194	0.5	3.88E-06	7.56E-03	ACSS2	-	-	-
8	7	rs80870743	112,227	0.496	1.24E-06	5.12E-03	ACSS2	-	-	-
8	7	rs81396301	113,410	0.496	1.52E-06	5.12E-03	ACSS2	-	-	-
9	18	rs81467823	23,933	0.122	4.39E-06	7.56E-03	ACSS2	intron variant	ENSSSCG00000016608	IQUB
9	18	rs81467842	24,097	0.122	4.39E-06	7.56E-03	ACSS2	-	-	-
9	18	rs81467847	24,134	0.122	4.39E-06	7.56E-03	ACSS2	upstream gene variant	ENSSSCG00000016609	SLC13A1
9	18	rs81467850	24,151	0.122	4.39E-06	7.56E-03	ACSS2	intron variant	ENSSSCG00000016609	SLC13A1
10	6	rs81395771	45,342	0.563	1.05E-06	3.81E-02	HIF1AN	upstream gene variant	ENSSSCG00000002927	LRFN3
10	6	rs81268228	47,685	0.366	6.42E-06	3.87E-02	HIF1AN	intron variant	ENSSSCG00000026794	SIRT2
10	6	rs81336138	47,699	0.634	6.42E-06	3.87E-02	HIF1AN	intron variant	ENSSSCG00000026794	SIRT2
10	6	rs81325240	47,715	0.366	6.42E-06	3.87E-02	HIF1AN	downstream gene variant	ENSSSCG00000002972	-
11	9	rs81409378	39,864	0.399	5.39E-06	3.87E-02	HIF1AN	intron variant	ENSSSCG00000022445	TEX12
11	9	rs81409385	40,174	0.504	8.86E-06	4.58E-02	HIF1AN	-	-	-
11	9	rs81283943	40,742	0.63	5.52E-06	3.87E-02	HIF1AN	intron variant	ENSSSCG00000015045	NCAM1
12	12	rs81309148	54,901	0.336	2.26E-06	1.51E-02	DGAT2	intron variant	ENSSSCG00000039056	-
12	12	rs81262159	56,062	0.546	2.10E-05	5.07E-02	DGAT2	intron variant	ENSSSCG00000018015	DNAH9
12	12	rs81312749	60,317	0.874	8.43E-06	2.77E-02	DGAT2	intron variant	ENSSSCG00000018047	FAM83G
12	12	rs81322820	60,352	0.845	2.06E-05	5.07E-02	DGAT2	intron variant	ENSSSCG00000018049	SLC5A10
Inter-	Chr	SNP	Position	MAF	p-value	FDR	Associated	Consequence	Ensembl GeneID	Gene

val			(Mb)				Gene			Symbol
13	16	rs81458250	1,094	0.454	1.50E-06	1.35E-02	DGAT2	intron variant	ENSSSCG00000016780	CTNND2
13	16	rs81459294	1,133	0.538	1.88E-05	5.07E-02	DGAT2	intron variant	ENSSSCG00000016780	CTNND2
13	16	rs81283619	1,305	0.433	1.47E-06	1.35E-02	DGAT2	intron variant	ENSSSCG00000016780	CTNND2
13	16	rs81464516	1,428	0.584	4.74E-06	2.07E-02	DGAT2	intron variant	ENSSSCG00000016780	CTNND2
13	16	rs81457374	2,779	0.248	2.51E-06	1.51E-02	DGAT2	-	-	-
13	16	rs81262166	3,001	0.42	1.77E-07	6.40E-03	DGAT2	-	-	-
13	16	rs81312810	3,422	0.525	8.58E-07	1.35E-02	DGAT2	-	-	-
13	16	rs81458107	3,850	0.538	1.83E-05	5.07E-02	DGAT2	intron variant	ENSSSCG00000016781	TRIO
13	16	rs81328340	5,131	0.122	5.72E-06	2.07E-02	DGAT2	-	-	-
13	16	rs81283415	5,188	0.895	5.18E-06	2.07E-02	DGAT2	intron variant	ENSSSCG00000028239	FBXL7
13	16	rs81459069	5,258	0.895	5.18E-06	2.07E-02	DGAT2	-	-	-
14	2	rs81341288	70	0.912	1.12E-07	1.76E-04	IGF2	downstream gene variant	ENSSSCG00000014559	PSMD13
14	2	rs81306755	145	0.218	8.71E-26	5.25E-22	IGF2	intron variant	ENSSSCG00000014565	-
14	2	rs81328276	236	0.782	8.71E-26	5.25E-22	IGF2	intron variant	ENSSSCG00000024569	ANO9
14	2	rs81317307	310	0.218	8.71E-26	5.25E-22	IGF2	intron variant	ENSSSCG00000027045	LRRC56
14	2	rs81339115	422	0.782	8.71E-26	5.25E-22	IGF2	intron variant	ENSSSCG00000012850	DEAF1
14	2	rs81341763	677	0.782	8.71E-26	5.25E-22	IGF2	-	-	-
14	2	IGF2	1,483	0.223	2.27E-28	8.21E-24	IGF2	-	-	-
14	2	rs81343851	2,149	0.58	8.60E-07	1.00E-03	IGF2	intron variant	ENSSSCG00000012857	CARS
14	2	rs81333729	2,380	0.542	2.60E-07	3.76E-04	IGF2	intron variant	ENSSSCG00000021181	DHCR7
14	2	rs81318741	2,594	0.87	1.93E-06	1.83E-03	IGF2	intron variant	ENSSSCG00000033043	SHANK2
14	2	rs81346169	2,694	0.504	8.00E-07	9.64E-04	IGF2	intron variant	ENSSSCG00000033043	SHANK2
14	2	rs81330032	2,878	0.046	3.31E-05	2.30E-02	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81252426	2,984	0.487	1.08E-06	1.08E-03	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81330112	3,058	0.487	1.08E-06	1.08E-03	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81336288	3,062	0.487	1.08E-06	1.08E-03	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81341267	3,094	0.487	1.08E-06	1.08E-03	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs333411238	3,257	0.416	5.75E-05	3.49E-02	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81322199	3,689	0.084	9.77E-08	1.61E-04	IGF2	intron variant	ENSSSCG00000031191	-
14	2	rs81356987	3,859	0.298	1.23E-17	5.56E-14	IGF2	-	-	-
14	2	rs341817021	3,978	0.815	2.78E-12	8.38E-09	IGF2	-	-	-
14	2	rs81357266	3,985	0.269	4.16E-21	2.15E-17	IGF2	-	-	-
14	2	rs81364067	4,412	0.521	6.78E-06	5.57E-03	IGF2	intron variant	ENSSSCG00000012884	PPP6R3
14	2	rs81237341	4,671	0.088	3.96E-05	2.70E-02	IGF2	-	-	-
14	2	rs81368683	4,966	0.567	6.53E-09	1.31E-05	IGF2	upstream gene variant	ENSSSCG00000012896	NDUFV1
14	2	rs81294446	5,339	0.42	1.86E-08	3.36E-05	IGF2	intron variant	ENSSSCG00000029637	KDM2A
14	2	rs81346312	5,372	0.147	3.37E-06	2.90E-03	IGF2	-	-	-
14	2	rs81326091	5,417	0.046	4.51E-05	2.97E-02	IGF2	intron variant	ENSSSCG00000024837	SYT12
14	2	rs81356796	5,827	0.424	5.34E-07	7.16E-04	IGF2	intron variant	ENSSSCG00000012934	CCS
14	2	rs81285769	5,830	0.424	5.34E-07	7.16E-04	IGF2	upstream gene variant	ENSSSCG00000012933	CCDC87
14	2	rs81357172	6,263	0.139	2.37E-05	1.75E-02	IGF2	-	-	-
14	2	rs81213587	6,922	0.408	1.70E-07	2.56E-04	IGF2	synonymous variant	ENSSSCG00000012997	-
14	2	rs81358499	6,949	0.651	3.66E-08	6.30E-05	IGF2	upstream gene variant	ENSSSCG00000012997	-
14	2	rs81358530	6,962	0.466	3.22E-09	7.27E-06	IGF2	intron variant	ENSSSCG00000012999	CAPN1
14	2	rs81359193	7,493	0.231	2.18E-06	1.97E-03	IGF2	intron variant	ENSSSCG00000033188	-
14	2	rs81359364	7,778	0.643	3.26E-05	2.30E-02	IGF2	upstream gene variant	ENSSSCG00000013042	-
14	2	rs81359337	7,852	0.429	1.66E-13	6.01E-10	IGF2	intron variant	ENSSSCG00000034755	-
14	2	rs81359542	7,987	0.643	7.75E-07	9.64E-04	IGF2	intron variant	ENSSSCG00000013043	MACROD1
14	2	rs81359616	8,060	0.597	8.97E-13	2.95E-09	IGF2	intron variant	ENSSSCG00000013043	MACROD1
Inter-	**Chr**	**SNP**	**Position**	**MAF**	**p-value**	**FDR**	**Associated**	**Consequence**	**Ensembl GeneID**	**Gene**

val			(Mb)				Gene			Symbol
14	2	rs81359810	8,474	0.87	1.82E-06	1.78E-03	IGF2	intron variant	ENSSSCG00000013056	LGALS12
14	2	rs81360111	8,647	0.761	9.97E-14	4.01E-10	IGF2	intron variant	ENSSSCG00000028537	-
14	2	rs81359966	8,687	0.777	4.37E-06	3.67E-03	IGF2	intron variant	ENSSSCG00000028537	-
14	2	rs81359986	8,713	0.592	8.29E-12	2.31E-08	IGF2	intron variant	ENSSSCG00000028537	-
14	2	rs81360021	8,765	0.357	4.05E-11	1.04E-07	IGF2	-	-	-
14	2	rs81330355	8,819	0.714	7.51E-07	9.64E-04	IGF2	downstream gene variant	ENSSSCG00000023571	SLC22A6
14	2	rs81270678	8,882	0.483	6.95E-05	4.12E-02	IGF2	-	-	-
14	2	rs81312355	8,894	0.403	1.01E-06	1.08E-03	IGF2	synonymous variant	ENSSSCG00000022404	SLC3A2
14	2	rs81214179	8,936	0.5	2.42E-05	1.75E-02	IGF2	upstream gene variant	ENSSSCG00000026293	STX5
14	2	rs81474834	9,252	0.307	2.15E-06	1.97E-03	IGF2	intron variant	ENSSSCG00000036669	-
14	2	rs338641431	9,757	0.845	1.84E-05	1.39E-02	IGF2	downstream gene variant	ENSSSCG00000039481	-
14	2	rs81360570	9,772	0.845	1.84E-05	1.39E-02	IGF2	intron variant	ENSSSCG00000013078	MYRF
14	2	rs81360547	9,787	0.857	4.71E-09	1.00E-05	IGF2	intron variant	ENSSSCG00000013078	MYRF
14	2	rs81296147	11,398	0.294	4.73E-05	3.05E-02	IGF2	downstream gene variant	ENSSSCG00000034506	MS4A2
14	2	rs81361529	11,420	0.178	2.34E-06	2.07E-03	IGF2	5 prime UTR variant	ENSSSCG00000034506	MS4A2
14	2	rs81474931	11,764	0.622	1.73E-09	4.16E-06	IGF2	non coding transcript exon variant	ENSSSCG00000037095	-
14	2	rs81238148	11,975	0.601	1.57E-05	1.23E-02	IGF2	upstream gene variant	ENSSSCG00000031679	-
14	2	rs81362332	12,168	0.391	4.10E-05	2.74E-02	IGF2	-	-	-
14	2	rs81362513	12,438	0.324	1.11E-08	2.11E-05	IGF2	intron variant	ENSSSCG00000032136	-
14	2	rs81323907	12,604	0.412	9.26E-06	7.44E-03	IGF2	missense variant	ENSSSCG00000013157	OR5B21
14	2	rs81474655	16,610	0.92	5.69E-05	3.49E-02	IGF2	intron variant	ENSSSCG00000013270	CRY2
14	2	rs81315092	17,672	0.231	5.78E-05	3.49E-02	IGF2	upstream gene variant	ENSSSCG00000013278	TSPAN18
14	2	rs81368350	17,683	0.769	5.78E-05	3.49E-02	IGF2	-	-	-
15	2	rs81312054	128,074	0.004	3.65E-06	2.64E-02	LXRA	intron variant	ENSSSCG00000014242	ZNF608
15	2	rs81364467	128,106	0.996	3.65E-06	2.64E-02	LXRA	intron variant	ENSSSCG00000014242	ZNF608
15	2	rs81364486	128,171	0.996	3.65E-06	2.64E-02	LXRA	-	-	-
16	7	rs80910844	100,110	0.008	2.81E-05	3.37E-02	PPARG	-	-	-
16	7	rs80947173	106,923	0.992	2.81E-05	3.37E-02	PPARG	-	-	-
16	7	rs80875484	107,365	0.008	2.81E-05	3.37E-02	PPARG	-	-	-
16	7	rs80830536	108,051	0.992	2.81E-05	3.37E-02	PPARG	-	-	-
16	7	rs80950513	108,219	0.992	2.81E-05	3.37E-02	PPARG	-	-	-
17	10	rs81222321	20,791	0.008	4.16E-09	2.98E-05	PPARG	-	-	-
17	10	rs81422104	21,752	0.008	4.16E-09	2.98E-05	PPARG	-	-	-
17	10	rs81244901	27,849	0.038	7.94E-07	1.69E-03	PPARG	intron variant	ENSSSCG00000010949	DAPK1
17	10	rs81329765	27,894	0.954	3.06E-05	3.37E-02	PPARG	intron variant	ENSSSCG00000010949	DAPK1
18	14	rs80955078	73,076	0.979	5.73E-09	2.98E-05	PPARG	downstream gene variant	ENSSSCG00000010259	TYSND1
18	14	rs80897247	73,237	0.979	5.73E-09	2.98E-05	PPARG	intron variant	ENSSSCG00000010267	LRRC20
18	14	rs80892145	73,289	0.979	5.73E-09	2.98E-05	PPARG	intron variant	ENSSSCG00000010267	LRRC20
19	16	rs81283619	1,305	0.433	3.45E-05	3.46E-02	PPARG	intron variant	ENSSSCG00000016780	CTNND2
19	16	rs81464516	1,428	0.584	4.11E-05	3.91E-02	PPARG	intron variant	ENSSSCG00000016780	CTNND2
19	16	rs81328340	5,131	0.122	2.16E-05	3.37E-02	PPARG	-	-	-
19	16	rs81283415	5,188	0.895	1.66E-05	2.73E-02	PPARG	intron variant	ENSSSCG00000028239	FBXL7
19	16	rs81459069	5,258	0.895	1.66E-05	2.73E-02	PPARG	-	-	-
20	17	rs81245673	29,246	0.021	1.63E-07	4.53E-04	PPARG	intron variant	ENSSSCG00000007108	KIZ
Inter-	**Chr**	**SNP**	**Position**	**MAF**	**p-value**	**FDR**	**Associated**	**Consequence**	**Ensembl GeneID**	**Gene**

val			(Mb)				Gene			Symbol
20	17	rs81322848	34,374	0.946	2.99E-05	3.37E-02	PPARG	intron variant	ENSSSCG00000007198	ANGPT4
20	17	rs80925918	34,516	0.029	3.54E-07	7.99E-04	PPARG	-	-	-
20	17	rs81466341	34,809	0.021	9.90E-09	2.98E-05	PPARG	3 prime UTR variant	ENSSSCG00000039862	TRIB3
20	17	rs80805843	34,820	0.021	9.90E-09	2.98E-05	PPARG	upstream gene variant	ENSSSCG00000039862	TRIB3
20	17	rs335077105	36,613	0.979	9.90E-09	2.98E-05	PPARG	synonymous variant	ENSSSCG00000029295	BPIFB4
20	17	rs340815635	36,637	0.021	9.90E-09	2.98E-05	PPARG	missense variant	ENSSSCG00000029295	BPIFB4
20	17	rs80903619	40,658	0.025	3.53E-07	7.99E-04	PPARG	-	-	-
20	17	rs80832461	41,811	0.975	3.53E-07	7.99E-04	PPARG	intron variant	ENSSSCG00000007350	PPP1R16B
21	2	rs81252691	38,875	0.004	1.79E-10	6.48E-06	PPARGC1A	-	-	-
21	2	rs81329722	39,427	0.021	1.94E-05	4.29E-02	PPARGC1A	intron variant	ENSSSCG00000013351	NAV2
21	2	rs81330475	39,626	0.979	1.94E-05	4.29E-02	PPARGC1A	intron variant	ENSSSCG00000013351	NAV2
22	6	rs81318862	82,375	0.029	4.13E-07	6.15E-03	PPARGC1A	-	-	-
22	6	rs81389632	89,786	0.036	1.05E-05	3.79E-02	PPARGC1A	intron variant	ENSSSCG00000003622	CSMD2
22	6	rs81389723	94,163	0.029	5.11E-07	6.15E-03	PPARGC1A	-	-	-
23	7	rs80910844	100,110	0.008	2.73E-05	4.29E-02	PPARGC1A	-	-	-
23	7	rs80947173	106,923	0.992	2.73E-05	4.29E-02	PPARGC1A	-	-	-
23	7	rs80875484	107,365	0.008	2.73E-05	4.29E-02	PPARGC1A	-	-	-
23	7	rs80830536	108,051	0.992	2.73E-05	4.29E-02	PPARGC1A	-	-	-
23	7	rs80950513	108,219	0.992	2.73E-05	4.29E-02	PPARGC1A	-	-	-
24	17	rs81245673	29,246	0.021	1.35E-05	4.29E-02	PPARGC1A	intron variant	ENSSSCG00000007108	KIZ
24	17	rs81466301	34,257	0.021	3.38E-06	1.36E-02	PPARGC1A	-	-	-
24	17	rs81466341	34,809	0.021	3.38E-06	1.36E-02	PPARGC1A	3 prime UTR variant	ENSSSCG00000039862	TRIB3
24	17	rs80805843	34,820	0.021	3.38E-06	1.36E-02	PPARGC1A	upstream gene variant	ENSSSCG00000039862	TRIB3
24	17	rs335077105	36,613	0.979	3.38E-06	1.36E-02	PPARGC1A	synonymous variant	ENSSSCG00000029295	BPIFB4
24	17	rs340815635	36,637	0.021	3.38E-06	1.36E-02	PPARGC1A	missense variant	ENSSSCG00000029295	BPIFB4
24	17	rs81466504	39,208	0.038	3.39E-06	1.36E-02	PPARGC1A	intron variant	ENSSSCG00000007311	PHF20
24	17	rs80903619	40,658	0.025	1.80E-05	4.29E-02	PPARGC1A	-	-	-
24	17	rs80832461	41,811	0.975	1.80E-05	4.29E-02	PPARGC1A	intron variant	ENSSSCG00000007350	PPP1R16B
25	17	rs81466301	34,257	0.021	1.39E-05	3.35E-02	SCD	-	-	-
25	17	rs80925918	34,516	0.029	3.22E-06	3.35E-02	SCD	-	-	-
25	17	rs81466341	34,809	0.021	1.39E-05	3.35E-02	SCD	3 prime UTR variant	ENSSSCG00000039862	TRIB3
25	17	rs80805843	34,820	0.021	1.39E-05	3.35E-02	SCD	upstream gene variant	ENSSSCG00000039862	TRIB3
25	17	rs335077105	36,613	0.979	1.39E-05	3.35E-02	SCD	synonymous variant	ENSSSCG00000029295	BPIFB4
25	17	rs340815635	36,637	0.021	1.39E-05	3.35E-02	SCD	missense variant	ENSSSCG00000029295	BPIFB4

Table S3: Significant eQTLs found for the 45-muscle gene expression study in each backcross independently. Start and end positions refer to the eQTL interval and are based on Sscrofa 11.1 assembly. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval. For the *cis*-eQTLs regions only the analysed gene was annotated as positional candidate gene.

Interval	Gene	Chr	Start Position (bp)	End Position (bp)	Size (bp)	SNPs	Type of eQTL	Candidate genes
1_LD	*ACSM5*	1	135,745,082	139,353,656	3,608,574	3	*trans*	
2_LD	*ACSM5*	1	251,939,764	258,564,948	6,625,184	3	*trans*	
3_LD	*ACSM5*	3	16,219,274	48,604,309	32,385,035	80	*cis*	
4_LD	*ACSM5*	6	20,167,972	32,407,613	12,239,641	3	*trans*	*E2F4* and *NFATC3*
5_LD	*ACSM5*	8	19,994,764	20,481,624	486.860	4	*trans*	
6_LD	*ACSM5*	10	60,180,539	60,634,967	454.428	4	*trans*	
7_LD	*HIF1AN*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
8_LD	*HIF1AN*	5	86,771,075	93,003,178	6,232,103	3	*trans*	
9_LD	*HIF1AN*	7	62,413,886	65,963,698	3,549,812	16	*trans*	*NFKBIA*
10_LD	*MLXIPL*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
11_LD	*MLXIPL*	9	18,204,538	26,457,702	8,253,164	8	*trans*	
12_LD	*MLXIPL*	13	81,152,914	81,639,287	486.373	8	*trans*	
13_LD	*CREG1*	2	119,937,944	136,199,281	16,261,337	18	*trans*	*TCF7*
14_LD	*DGAT2*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
15_LD	*DGAT2*	7	62,413,886	65,963,698	3,549,812	17	*trans*	*NFKBIA*
16_LD	*DGAT2*	9	106,297,336	123,003,538	16,706,202	6	*trans*	
17_LD	*FOS*	11	8,855,571	10,399,072	1,543,501	3	*trans*	
18_LD	*IGF2*	2	1,000,000	11,764,773	10,764,773	28	*cis*	
19_LD	*MGLL*	9	54,051,728	55,857,335	1,805,607	6	*trans*	
20_LD	*MGLL*	13	47,034,430	80,937,093	33,902,663	127	*cis*	*TADA3* and *PPARG*
21_LD	*MGLL*	13	107,655,572	109,659,033	2,003,461	4	*trans*	
22_LD	*NCOA2*	4	64,491,574	69,900,791	5,409,217	4	*cis*	
23_LD	*PIK3R1*	16	63,606,393	63,825,864	219.471	3	*trans*	*EBF1*
24_LD	*PPARG*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
25_LD	*PPARGC1A*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
26_LD	*SCD*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
1_DU	*ACSM5*	1	13,163,823	14,456,261	1,292,438	3	*trans*	*ESR1*
2_DU	*ACSM5*	1	246,699,639	258,930,988	12,231,349	4	*trans*	
3_DU	*ACSM5*	2	8,713,993	26,475,888	17,761,895	10	*trans*	*ESRRA*
4_DU	*ACSM5*	3	4,163,933	32,847,659	28,683,726	46	*cis*	
5_DU	*ACSM5*	4	9,484,620	18,736,459	9,251,839	12	*trans*	
6_DU	*ACSM5*	6	2,754,836	7,566,303	4,811,467	19	*trans*	
7_DU	*ACSM5*	7	42,263,721	46,283,437	4,019,716	5	*trans*	
8_DU	*ACSM5*	11	22,645,575	31,980,529	9,334,954	4	*trans*	
9_DU	*ACSM5*	11	44,662,760	76,803,907	32,141,147	27	*trans*	

Interval	Gene	Chr	Start Position (bp)	End Position (bp)	Size (bp)	SNPs	Type of eQTL	Candidate genes
10_DU	*ACSM5*	12	59,035,392	59,945,246	909.854	5	*trans*	*SREBF1* and *NCOR1*
11_DU	*ACSM5*	13	175,321,576	191,415,117	16,093,541	12	*trans*	
12_DU	*ACSM5*	18	5,518,202	55,272,842	49,754,640	33	*trans*	*CREB5* and *MTPN*
13_DU	*ACAA2*	1	180,853,163	180,902,764	49.601	3	*trans*	
14_DU	*ACAA2*	1	197,608,882	203,632,016	6,023,134	11	*trans*	*PLIN2*
15_DU	*CREG1*	1	197,608,882	197,974,601	365.719	7	*trans*	
16_DU	*DGAT2*	2	17,672,685	26,400,002	8,727,317	7	*trans*	
17_DU	*ETS1*	9	34,248,222	44,486,848	10,238,626	9	*trans*	
18_DU	*IGF2*	2	1,000,000	17,640,519	16,640,519	107	*cis*	*NR1H3*
19_DU	*IGF2*	4	107,415,236	110,758,451	3,343,215	6	*trans*	
20_DU	*LPIN1*	4	35,029,394	48,336,784	13,307,390	5	*trans*	
21_DU	*LPIN1*	4	67,460,687	84,642,594	17,181,907	7	*trans*	*RXRG*
22_DU	*LPIN1*	7	108,190,681	111,428,030	3,237,349	3	*trans*	
23_DU	*LPIN1*	15	103,751,874	104,093,912	342.038	6	*trans*	*AOX1*
24_DU	*NCOA1*	1	197,608,882	203,632,016	6,023,134	7	*trans*	*PLIN2*
25_DU	*NCOA6*	1	197,608,882	203,632,016	6,023,134	10	*trans*	*PLIN2*
26_DU	*PDHX*	1	159,785,087	167,989,461	8,204,374	6	*trans*	*SMAD3* and *PIAS1*
27_DU	*PDHX*	1	180,585,237	203,632,016	23,046,779	21	*trans*	*PLIN2, ESR2*
28_DU	*PPARA*	2	43,500,475	43,686,936	186.461	3	*trans*	
29_DU	*PPARA*	15	103,751,874	104,093,912	342.038	6	*trans*	*AOX1*
30_DU	*PPARA*	17	29,976,476	34,461,408	4,484,932	5	*trans*	*FOXA2*
31_DU	*PRKAA1*	1	197,608,882	197,974,601	365.719	7	*trans*	
32_DU	*PXMP3*	4	2,524,589	2,905,284	380.695	4	*trans*	
1_PI	*ACSM5*	1	271,398,918	271,933,557	534.639	8	*trans*	
2_PI	*ACSM5*	3	22,810,472	25,651,284	2,840,812	7	*cis*	
3_PI	*ACSM5*	3	64,193,306	67,608,274	3,414,968	4	*trans*	
4_PI	*ACSM5*	8	103,957,628	116,677,381	12,719,753	29	*trans*	
5_PI	*ACSM5*	12	27,551,282	31,648,894	4,097,612	3	*trans*	
6_PI	*ACSM5*	14	14,101,235	17,897,384	3,796,149	6	*trans*	
7_PI	*ACSM5*	16	62,156,018	71,272,174	9,116,156	4	*trans*	*EBF1*
8_PI	*ACSS2*	7	111,283,606	113,410,786	2,127,180	20	*trans*	
9_PI	*ACSS2*	18	23,933,285	24,151,664	218.379	4	*trans*	
10_PI	*HIF1AN*	6	45,342,655	47,715,812	2,373,157	4	*trans*	
11_PI	*HIF1AN*	9	39,864,484	40,742,876	878.392	3	*trans*	
12_PI	*DGAT2*	12	54,901,488	60,352,653	5,451,165	4	*trans*	*SREBF1*
13_PI	*DGAT2*	16	1,094,075	5,258,475	4,164,400	11	*trans*	
14_PI	*IGF2*	2	1,000,000	17,683,291	16,683,291	61	*cis/trans*	*NR1H3*
15_PI	*LXRA*	2	128,074,209	128,171,493	97.284	3	*trans*	
16_PI	*PPARG*	7	100,110,224	108,219,850	8,109,626	5	*trans*	*DIO2*
17_PI	*PPARG*	10	20,791,457	27,894,378	7,102,921	4	*trans*	
18_PI	*PPARG*	14	73,076,159	73,289,391	213.232	3	*trans*	
19_PI	*PPARG*	16	1,305,017	5,258,475	3,953,458	5	*trans*	

Interval	Gene	Chr	Start Position (bp)	End Position (bp)	Size (bp)	SNPs	Type of eQTL	Candidate genes
20_PI	*PPARG*	17	29,246,810	41,811,151	12,564,341	10	*trans*	*RBL1, E2F1* and *FOXA2*
21_PI	*PPARGC1A*	2	38,875,533	39,626,373	750.840	3	*trans*	
22_PI	*PPARGC1A*	6	82,375,634	94,163,280	11,787,646	3	*trans*	
23_PI	*PPARGC1A*	7	100,110,224	108,219,850	8,109,626	5	*trans*	*DIO2*
24_PI	*PPARGC1A*	17	29,246,810	41,811,151	12,564,341	9	*trans*	*RBL1, E2F1* and *FOXA2*
25_PI	*SCD*	17	34,257,472	36,637,601	2,380,129	6	*trans*	*E2F1*

Table S4: Significant *trans*-eQTLs for the hotspot regions found in each backcross independently. Start and end positions refer to the eQTL interval and are based on Sscrofa 11.1 assembly. Gene annotation was performed considering one additional Mb at the start and at the end of the eQTL interval. SNPs column indicates the number of SNPs within the eQTL interval.

Interval	Gene	Chr	Start Position (bp)	End Position (bp)	Size (bp)	SNPs	Type of eQTL	Candidate genes
15_LD	*DGAT2*	7	62,413,886	65,963,698	3,549,812	17	*trans*	*NFKBIA*
9_LD	*HIF1AN*	7	62,413,886	65,963,698	3,549,812	16	*trans*	*NFKBIA*
7_LD	*HIF1AN*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
10_LD	*MLXIPL*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
13_LD	*CREG1*	2	119,937,944	136,199,281	16,261,337	18	*trans*	*TCF7*
14_LD	*DGAT2*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
24_LD	*PPARG*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
25_LD	*PPARGC1A*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
26_LD	*SCD*	2	119,937,944	128,693,964	8,756,020	12	*trans*	
14_DU	*ACAA2*	1	197,608,882	203,632,016	6,023,134	11	*trans*	*PLIN2*
15_DU	*CREG1*	1	197,608,882	197,974,601	365.719	7	*trans*	
24_DU	*NCOA1*	1	197,608,882	203,632,016	6,023,134	7	*trans*	*PLIN2*
25_DU	*NCOA6*	1	197,608,882	203,632,016	6,023,134	10	*trans*	*PLIN2*
27_DU	*PDHX*	1	180,585,237	203,632,016	23,046,779	21	*trans*	*PLIN2, ESR2*
31_DU	*PRKAA1*	1	197,608,882	197,974,601	365.719	7	*trans*	
23_DU	*LPIN1*	15	103,751,874	104,093,912	342.038	6	*trans*	*AOX1*
29_DU	*PPARA*	15	103,751,874	104,093,912	342.038	6	*trans*	*AOX1*
16_PI	PPARG	7	100,110,224	108,219,850	8,109,626	5	*trans*	*DIO2*
23_PI	PPARGC1A	7	100,110,224	108,219,850	8,109,626	5	*trans*	*DIO2*
20_PI	PPARG	17	29,246,810	41,811,151	12,564,341	10	*trans*	*RBL1, E2F1* and *FOXA2*
24_PI	PPARGC1A	17	29,246,810	41,811,151	12,564,341	9	*trans*	*RBL1, E2F1* and *FOXA2*
25_PI	SCD	17	34,257,472	36,637,601	2,380,129	6	*trans*	*E2F1*

7.2. Supplementary material Paper II: 'Analysis of porcine IGF2 gene expression in adipose tissue and its effect on fatty acid composition'

Table S1. Primers used for *IGF2* gene expression quantification by RT-qPCR.

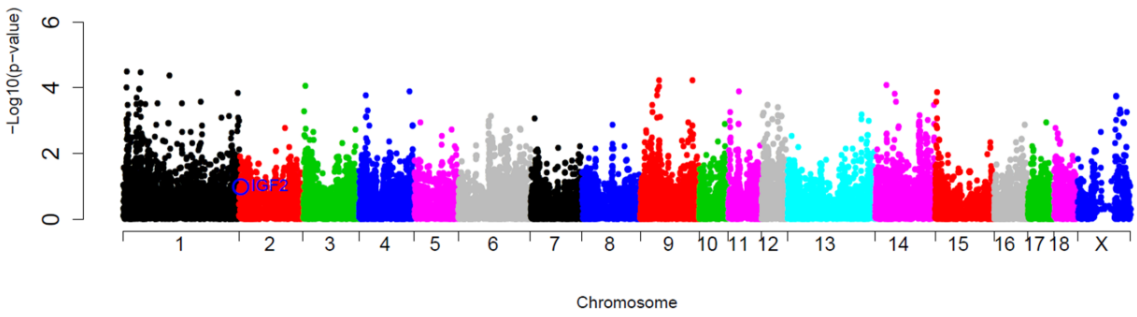
Primer Name	Primer Sequence	Location	Amplicon Lenght
IGF2_Fw[1]	5'-GACCGTGCTTCCGGACAA-3'	Exon 3 and 4	81 bp
IGF2_Rv	5'-CGTTGGGCGGACTGCTT-3'	Exon 4	
ACTB_Fw[2]	5'-CAAGGACCTCTACGCCAACAC-3'	Exon 5	81 bp
ACTB_Rv	5'-TGGAGGCGCGATGATCTT-3	Exon 5 and 6	
TBP_Fw[3]	5'-CAGAATGATCAAACCGAGAATTGT-3'	Exon 9	80 bp
TBP_Rv	5'-CTGCTCTGACTTTAGCACCTGTTAA-3'	Exon 9 and 10	

[1]Primers were designed from the GenBank X56094 sequence

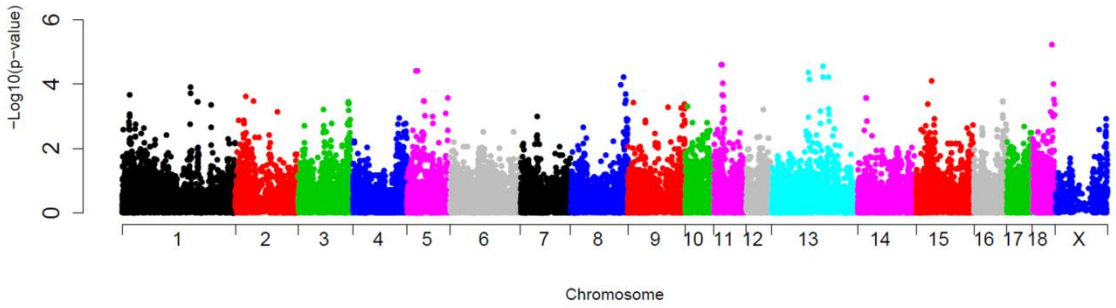
[2]Primers were designed from the GenBank NC_010445 sequence

[3]Primers were designed from the GenBank DQ845178 sequence

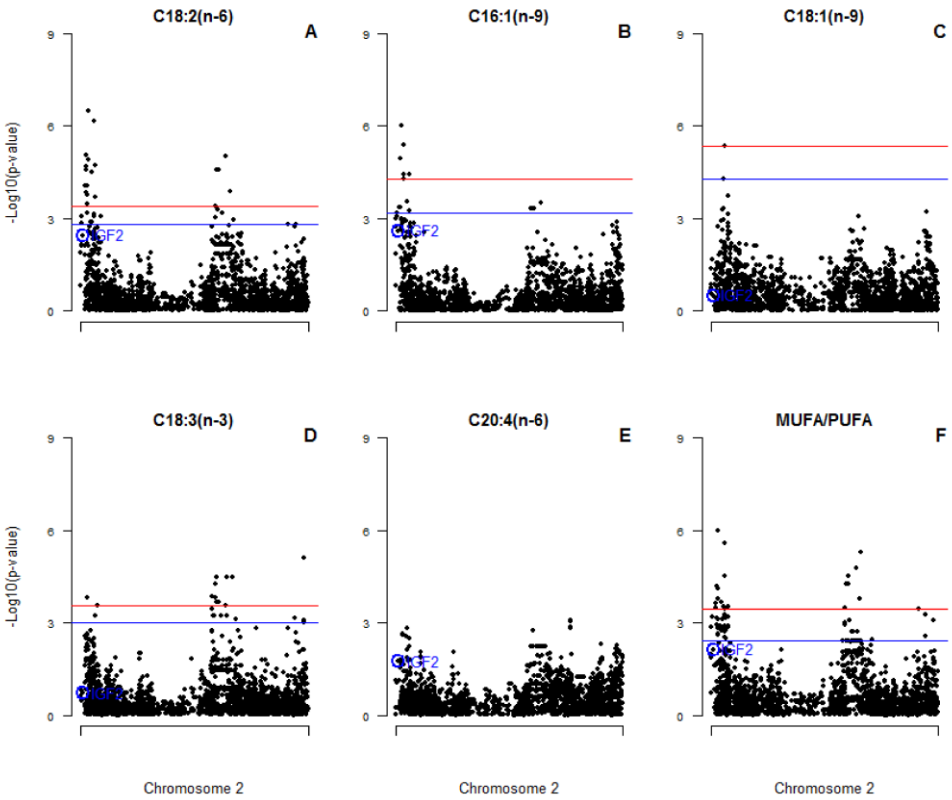
S1 Fig. GWAS plot of adipose tissue *IGF2* gene expression in BC1_LD. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.



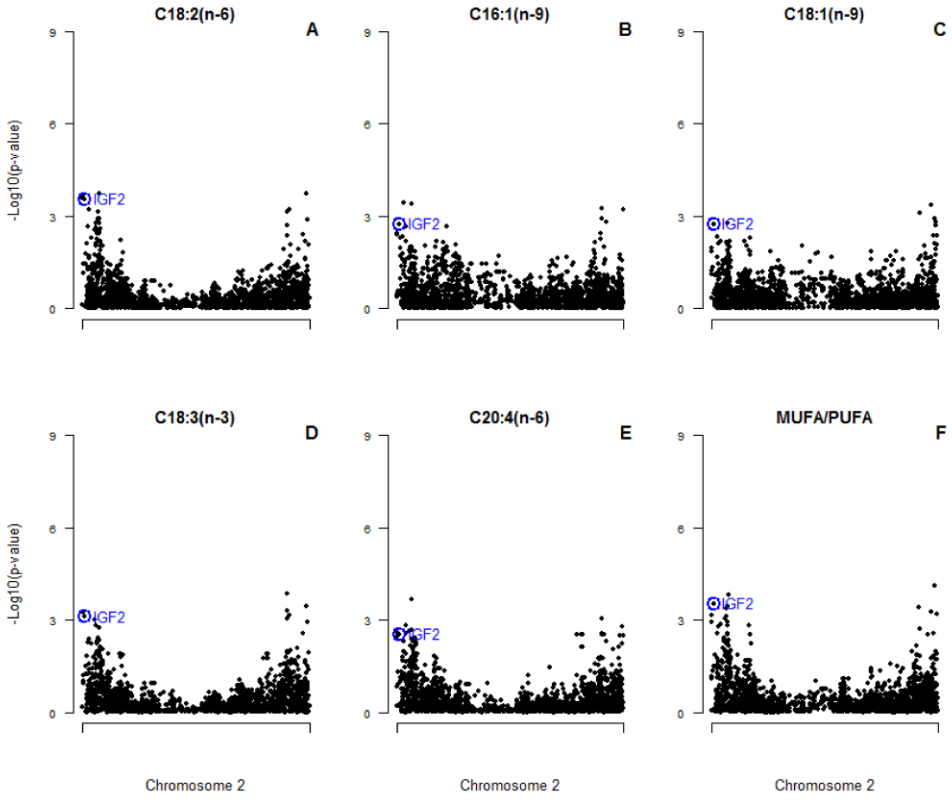
S2 Fig. GWAS plot of adipose tissue *IGF2* gene expression in BC1_DU using *IGF2:g.3072G>A* polymorphism as a fixed effect. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value <0.1).



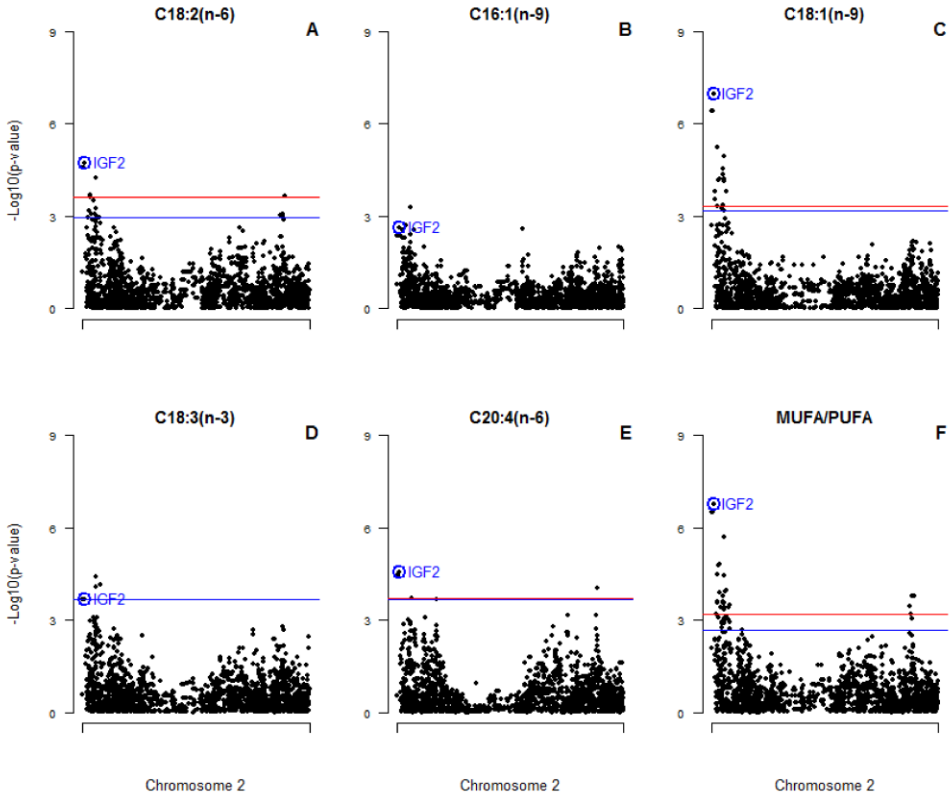
S3 Fig. Plot of SSC2 SNPs association for significant FAs in BC1_LD. (A) linoleic acid, (B) hexadecanoic acid, (C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The IGF2:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.



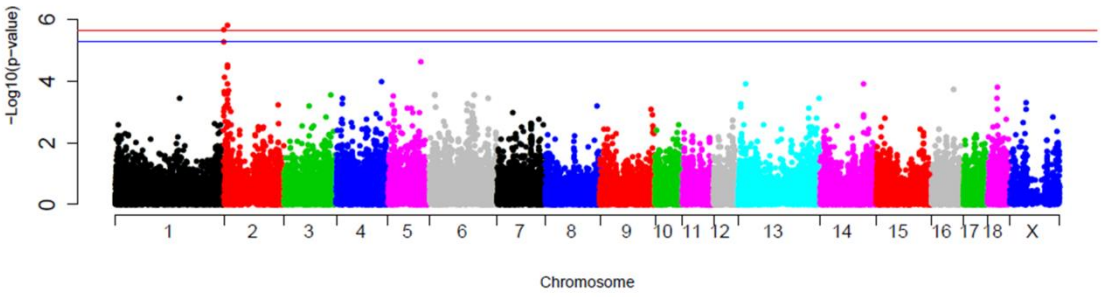
S4 Fig. Plot of SSC2 SNPs association for significant FAs in the BC1_DU. (A) linoleic acid, (B) hexadecanoic acid, (C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2*:*g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.



S5 Fig. Plot of SSC2 SNPs association for significant FAs in the BC1_PI. (A) linoleic acid, (B) hexadecanoic acid,(C) oleic acid, (D) α-linoleic acid, and (E) arachidonic acid, and (F) MUFA/PUFA ratio in adipose tissue in 3BCs. Chromosome 2 (SSC2) positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the chromosomal-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value < 0.1). The *IGF2:g.3072G>A* polymorphism is circled and labelled as IGF2 in colour blue.



S6 Fig. GWAS plot of BF thickness measure in the 3BCs animals. Chromosome positions in Mb based on *S. scrofa 11.1* assembly of the pig genome are represented in the X-axis and the –log10 (p-value) is on the Y-axis. The red horizontal line indicates the genome-wide significant level (FDR-based *q*-value < 0.05) and the blue horizontal line represents the genome-wide suggestive level (FDR-based *q*-value <0.1).



7.3. Supplementary material Paper III: 'Expression analysis of porcine miR-33a/b in liver, adipose tissue and *longissimus dorsi* muscle and its role in fatty acid metabolism'

Table S1. Primers used for *SREBF2*, *ACTB* and *TBP* gene expression quantification by RT-qPCR.

Primer Name	Primer Sequence	Location	Amplicon Lenght
SREBF2_Fw[1]	5'-GTACCGCTCCTCCATCAATGA-3'	Exon 5	87 bp
SREBF2_Rv	5'-AAAACACCAGACTTGTGCATCTTG-3'	Exon 5 and 6	
ACTB_Fw[2]	5'-CAAGGACCTCTACGCCAACAC-3'	Exon 5	81 bp
ACTB_Rv	5'-TGGAGGCGCGATGATCTT-3	Exon 5 and 6	
TBP_Fw[3]	5'-CAGAATGATCAAACCGAGAATTGT-3'	Exon 9	80 bp
TBP_Rv	5'-CTGCTCTGACTTTAGCACCTGTTAA-3'	Exon 9 and 10	

[1]Primers were designed from the GenBank XM_021091446.1 sequence

[2]Primers were designed from the GenBank NC_010445 sequence

[3]Primers were designed from the GenBank DQ845178 sequence

Table S2. List of genes analysed by RT-qPCR in each tissue (Puig-Oliveras *et al.*, 2016; Ballester, Ramayo-Caldas, *et al.*, 2017; Revilla *et al.*, 2018).

Gene_Name	Liver	Adipose tissue	Muscle	Common
ABCG8	X			
ACAA2			X	
ACSM5	X	X	X	X
ACSS1			X	
ACSS2			X	
ADIPOQ		X		
AGPAT2	X	X		
ALB			X	
ANGPT1			X	
ANK2		X		
APOA2	X			
APOB	X			
AQP7			X	
ARNT	X	X		
ATF3			X	
ATGL		X		

Gene_Name	Liver	Adipose tissue	Muscle	Common
CD36		X		
ChREBP	X	X	X	X
CPT1A	X	X		
CPT1b			X	
CREG1			X	
CROT	X	X	X	X
CYP2U1	X	X		
CYP7A1	X			
DGAT1	X	X	X	X
DGAT2		X	X	
EGF	X	X		
ELF1			X	
ELOVL5	X	X		
ELOVL6	X	X		
ESRRA	X	X		
ETS1			X	
FABP1	X			
FABP3			X	
FABP4		X		
FABP5	X	X	X	X
FADS1	X	X		
FADS2	X	X		
FADS3	X	X		
FATP1		X		
FATP4		X		
FOS			X	
HADH	X			
HIF1AN			X	
HNF3	X			
HNF4a	X			
HNF4g	X			
IGF2			X	
KLF10	X			
LIPC	X	X		
LPIN1	X	X	X	X
LXRa	X	X	X	X
MGLL		X	X	
MTTP	X			
NCOA1			X	
NCOA2			X	
NCOA6			X	
NFKB			X	

Gene_Name	Liver	Adipose tissue	Muscle	Common
NFKB1	X	X		
PDHX			X	
PIK3R1			X	
PLA2G12A	X	X	X	X
PLCB2	X	X		
PLIN5			X	
POU2F1	X	X		
PPAP2A	X	X	X	X
PPARA	X	X	X	X
PPARD	X	X	X	X
PPARg		X	X	
PPARGC1A	X		X	
PRKAA1			X	
PXMP3	X	X	X	X
RBP4		X		
RXRg		X	X	
SCAP	X	X		
SCD			X	
SCD1	X	X		
SETD7			X	
SLC2A4			X	
SP1			X	
SREBP1	X			
SREBP1c		X	X	
TBCK	X			
USF1	X	X		
TOTAL	**44**	**43**	**45**	**12**

Table S3. Pearson correlation values between miR-33a and *SREBF2* gene expression, and between miR-33b and *SREBF1* gene expression.

miR-33a – SREBF2		
Tissue	**r**	***p*-value**
Backfat	-0.07	6.57E-01
Muscle	-0.23	1.68E-01
Liver	0.50	1.12E-03
miR-33b – SREBF1		
Tissue	**r**	***p*-value**
Backfat	0.19	2.50E-01
Muscle	-0.12	4.68E-01
Liver	0.00	9.93E-01

7.4. Supplementary material additional study: *ELOVL6*

The ChIP protocol is outlined in Figure S.1., an is based on David *et al.*, 2017. However, several steps had been modified and adapted to the current study. For instance, in the chromatin preparation the cross-linking step was performed after the tissue disaggregation and the nuclei purification was performed by using a sucrose gradient, and then was checked in the fluorescence microscopy. Besides, cells were lysed, and we focused on the sonication step, which is considered crucial for the ChIP procedure because will determine the length for primer amplification in the ChIP-qPCR analysis. The optimal fragment length of sheared DNA is around 100-400 bp. We used a water bath Bioruptor (Diagenode) to test different cycles in order to obtain the higher amount of DNA fragments with the expected length, and sheared DNA was checked by gel electrophoresis. The immunoprecipitations were performed for ERα*, SREBF1* and *SP1*, the input (reference of the whole genomic content), the IgG as a negative control (to identify unspecific immunoprecipitated DNA) and both RNA-Pol II and H3 as positive controls. Finally, the reverse cross-linking was performed before DNA was extracted using phenol/chloroform method.

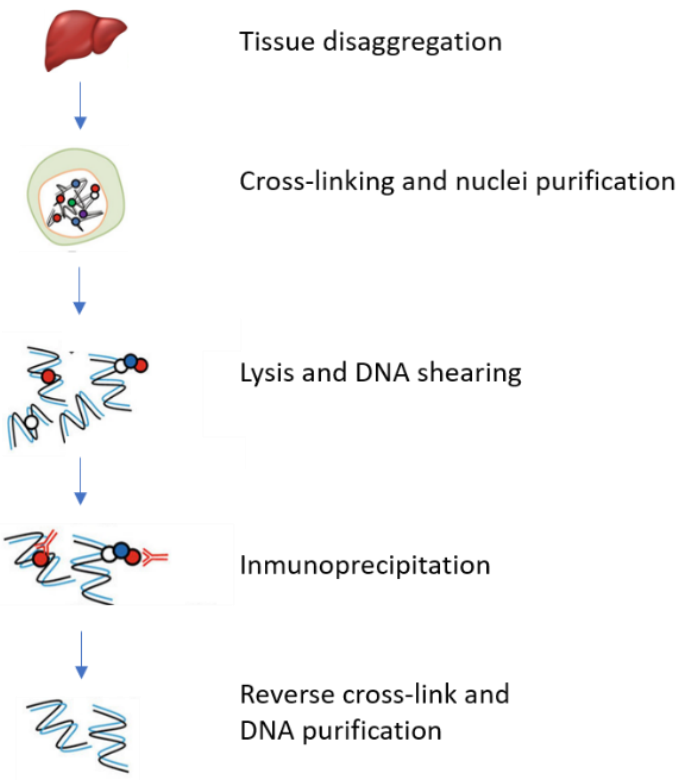


Figure S1. Chromatin immunoprecipitation (ChIP) protocol used in this thesis

After all, four animals carrying the *ELOVL6:c.-394G* allele and four animals carrying the *ELOVL6:c.-394A* allele were selected for the ChIP-qPCR analysis. We used *GAPDH* as a reference gene. Primers were designed using Primer3 and covering the region containing the *ELOVL6:c.-394G>A* polymorphism and the transcription factors binding sites. The ChIP-qPCR was performed for the immunoprecipitations corresponding to ERα, SREBF1 and SP1, the input, the IgG and both RNA-Pol II and H3 for all eight samples. Due to the low amount of DNA recovered, a high range of Ct values was obtained (Ct > 27), and in some cases no amplification was detected. Hence, we were not able to differentiate between samples with different genotypes and between samples and controls.

Acknowledgments