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

Anna Puig-Oliveras¹,², Yuliaxis Ramayo-Caldas³, Jordi Corominas¹,², Jordi Estelle³,⁴,⁵, Dafne Pérez-Montarelo⁶, Nicholas J. Hudson⁷, Joaquim Casellas²,⁸, Josep M. Folch¹,², Maria Ballester¹,²

¹ Departament de Genètica Animal, Centre de Recerca en Agrigenòmica (CRAG), Bellaterra, Spain, ² Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain, ³ Gènètica Animal i Biologia Integrativa UMR1313 (GABI), Institut National de la Recherche Agronomique (INRA), Jouy-en-Josas, France, ⁴ Gènètica Animal i Biología Integrativa UMR1313 (GABI), AgroParisTech, Jouy-en-Josas, France, ⁵ Laboratorio de Radiobiología y Etude du Génome (LREG), Commissariat à l’Énergie Atomique et aux énergies alternatives (CEA), Jouy-en-Josas, France, ⁶ Departamento de Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain, ⁷ Computational and Systems Biology, Commonwealth Scientific and Industrial Research Organisation (CSIRO) Animal, Food and Health Sciences QLD, Brisbane, Australia, ⁸ Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologies Agroalimentàries (IRTA), Lleida, Spain

Abstract

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

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

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


Editor: Hervé Guillou, INRA, France

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

Copyright: © 2014 Puig-Oliveras et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work has been funded by MICINN project AGL2011-29821-C02 (Ministerio de Economía y Competitividad) and by the Innovation Consolider-Ingenio 2010 Program (CSD2007-00036). A. Puig-Oliveras was funded by a Personal Investigador en Formación (PIF) PhD grant from the Universitat Autònoma de Barcelona (458-01-12011). J. Corominas was funded by a Formación de Personal Investigador (FPI) PhD grant from Spanish Ministerio de Educación (BES-2009-018223). Y. Ramayo by a Formación de Profesorado Universitario (FPU) PhD grant (AP2008-01450). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

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

Introduction

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

According to the Food and Agriculture Organization (FAO) [7], pork is the major source of meat intake by human, accounting for the 43% of the consumed meat worldwide. The taste and the quality of the cooked and the cured meat products depend on the oxidative stability of the muscle which is related to the fatty acid (FA) composition [8,9]. Furthermore, it is well known that genetic and environmental factors such as diet, are responsible for FA composition variation [10]. Besides its influence on meat taste, the FA composition in muscle has taken additional importance due to their nutritional value and human health-related benefits [11,12], particularly for its effects on human diseases such as cancers,
coronary heart diseases and atherosclerosis [11]. It has been reported that omega-3 FAs, such as α-linolenic acid (C18:3 n-3), are associated with the reduction of low density lipoprotein (LDL) cholesterol and blood triacylglycerols, as well as with the modulation of immune functions and inflammatory processes [13,14]. Artificial selection to increase meat production in pigs has caused a reduction of intramuscular fat (IMF) and changes in meat FA composition in some breeds. Hereby, there is an increasing interest in the pork industry on producing meat products with a higher IMF content and with a healthier FA profile, while maintaining a reduced amount of backfat [15].

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

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

Transcriptome analysis of swine muscle tissue
As described above, the Longissimus dorsi (LD) muscle transcriptome was sequenced in twelve sows (H = 6, L = 6) with extreme phenotypes for intramuscular FA composition. A total amount of 787.5 M of 75 bp paired-end reads was acquired from the RNA-Seq experiment. Sequence alignment was performed against the reference pig genome (Sus scrofa10.2) by using Tophat [19]. About 83.1% (76.5%–86.6%) of reads were mapped to the reference genome, of which 14.5% (12.4%–16.1%) did not map to unique genomic locations. A total of 85.1% (84.0%–87.6%) of the mapped reads correspond to annotated genes, 79.1% (77.5%–84.1%) of them were located in exons and 60.0% (3.6%–6.8%) in introns. The remaining 14.9% (12.4%–16.0%) of reads mapped to intergenic regions, indicating that they were not annotated in the reference genome (Table S1).

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

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

With the aim of improving the current porcine genome annotation, we took into account the intergenic transcripts identified with cufflinks (a mean of 4,440 transcripts) to determine whether these transcripts potentially coded for proteins. A total of 2,372 putative proteins were predicted by Augustus [22] corresponding to non-annotated transcripts of the Suscrofa10.2 genome assembly version. Among the 2,372 novel predicted proteins, only 1,406 (59.2%) had at least one orthologous gene identified with BLASTP option of Blast2GO, representing a total of 577 known genes (Table S4) [23]. These proteins corresponded to: 720 Sus scrofa, 17 Homo sapiens and 247 Bos taurus in silico predicted protein, and 476 Sus scrofa, 935 Homo sapiens, and 403 Bos taurus known proteins. The pig species was the only one showing a higher percentage of computationally predicted protein (60.2%) in comparison to known proteins (39.8%).

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

Differential gene expression analysis
A high correlation (r = 0.98, P-value <2.2×10−16) between H and L groups in the mean gene expression was found, showing that most of the genes had a similar behaviour. A total of 11,945 transcripts were used to perform the differential expression analysis after filtering. Using EdgeR program [24], 314 genes were identified as significantly differentially expressed between H and L groups. Whereas, employing DESeq [25], 208 genes were detected. Figure 1 shows the P-value distribution and how among the transcripts accepted as differentially expressed the selected cutoff of P-value <0.01 is clearly departing from the expected P-value (equivalent to a FDR ≤0.12).

A total of 131 genes (Table S5) overlapping in both analyses were selected as differentially expressed between H and L groups.
and, thereafter, used for pathway analysis (Figure 2). Fifty genes had a higher expression and 81 a lower expression in the H group (in comparison with L group). Remarkably, eighteen (C14H10orf116, ITPR2, TRIP10, BANF1, HIF1AN, CHAC1 and C14H10orf116) were also identified in a muscle microarray study of extreme animals for FA composition from the same backcross [27]. Other genes such as DNAJA4, ANKK1, MTH10 and TNFRSF12A were also common, but they were only detected by the DESeq program [23] in the RNA-Seq data.

**Functional analysis**

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

<table>
<thead>
<tr>
<th>Characters</th>
<th>H Mean</th>
<th>L Mean</th>
<th>Significance</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carcass weight (CW) (kg)</td>
<td>66.22±10.52</td>
<td>70.50±7.91</td>
<td>NS</td>
<td>4.44×10^{-1}</td>
</tr>
<tr>
<td>Ham weight (HW) (kg)</td>
<td>18.79±2.32</td>
<td>18.97±2.41</td>
<td>NS</td>
<td>8.96×10^{-1}</td>
</tr>
<tr>
<td>Shoulder weight (SW) (kg)</td>
<td>6.6±0.95</td>
<td>6.32±0.97</td>
<td>NS</td>
<td>6.16×10^{-1}</td>
</tr>
<tr>
<td>Intramuscular fat (IMF) (%)</td>
<td>1.94±0.65</td>
<td>1.69±0.64</td>
<td>NS</td>
<td>5.25×10^{-1}</td>
</tr>
</tbody>
</table>

**Fatty acids in intramuscular fat**

<table>
<thead>
<tr>
<th>Saturated FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myristic acid (C14:0)</td>
</tr>
<tr>
<td>Palmitic acid (C16:0)</td>
</tr>
<tr>
<td>Heptadecanoic acid (C17:0)</td>
</tr>
<tr>
<td>Stearic acid (C18:0)</td>
</tr>
<tr>
<td>Arachidic acid (C20:0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monounsaturated FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palmitoleic acid (C16:1 n-7)</td>
</tr>
<tr>
<td>Heptadecenoic acid (C17:1)</td>
</tr>
<tr>
<td>Oleic acid (C18:1 n-9)</td>
</tr>
<tr>
<td>Octadecenoic acid (C18:1 n-7)</td>
</tr>
<tr>
<td>Eicosanoic acid (C20:1 n-9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Polyunsaturated FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linoleic acid (C18:2 n-6)</td>
</tr>
<tr>
<td>α-Linolenic acid (C18:3 n-3)</td>
</tr>
<tr>
<td>Eicosadienoic acid (C20:2 n-6)</td>
</tr>
<tr>
<td>Eicosatrienoic acid (C20:3 n-6)</td>
</tr>
<tr>
<td>Arachidonic acid (C20:4 n-6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metabolic ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Chain Length (ACL)</td>
</tr>
<tr>
<td>Saturated FA (SFA)</td>
</tr>
<tr>
<td>Monounsaturated FA (MUFA)</td>
</tr>
<tr>
<td>Polyunsaturated FA (PUFA)</td>
</tr>
<tr>
<td>Peroxidability index (PI)</td>
</tr>
<tr>
<td>Double-bond index (DBI)</td>
</tr>
<tr>
<td>Unsaturated index (UI)</td>
</tr>
</tbody>
</table>
(7 genes, \(P\)-value = \(7.75 \times 10^{-7}\)) and endothelial nitric oxide synthase signalling (eNOS, 8 genes, \(P\)-value = \(2.34 \times 10^{-6}\)). On the other hand, using Babelomics we observed an overrepresentation of lipids and lipoproteins metabolism (6 genes, \(P\)-value = \(1.64 \times 10^{-5}\)), and also the peroxisome proliferator-activated receptors (PPAR, 4 genes, \(P\)-value = \(7.25 \times 10^{-7}\)) and the insulin (5 genes, \(P\)-value = \(1.00 \times 10^{-7}\)) signalling and the hemostasis (7 genes, \(P\)-value = \(1.64 \times 10^{-7}\)) pathway (Table 2).

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

Among the related specific functions for lipid metabolism, the top molecular functions identified with IPA were the oxidation (ACADVL, ACOX2, FABP3, PLIN1, PLIN5, PON2, SCD; \(P\)-value = \(3.59 \times 10^{-7}\)), accumulation (ACADVL, AQP7, FH, IDH1, PLIN1, PON2, RETSAT, SCD; \(P\)-value = \(7.90 \times 10^{-7}\)), synthesis

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

doi:10.1371/journal.pone.0099720.g001

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

doi:10.1371/journal.pone.0099720.g002
Table 2. Summary of the most significantly-overrepresented pathways of the differentially-expressed genes in muscle between High and Low groups for fatty acid composition traits.

<table>
<thead>
<tr>
<th>Babelomics</th>
<th>Category</th>
<th>Genes</th>
<th>P-value</th>
<th>IPA</th>
<th>Category</th>
<th>Genes</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism of Lipids and Lipoproteins</td>
<td>SCD, ACADVL, ACOX2, ID1, ID1, ALB</td>
<td>5.12 × 10^{-4}</td>
<td>Nitric Oxid Signalling in the Cardiovascular System</td>
<td>BDKB2, PIK3R3, PPKG1, PLN, ITPR2, PIK3R1, ATP2A2</td>
<td>7.75 × 10^{-7}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alanine, Aspartate and Glutamate Metabolism</td>
<td>ABAT, GOT1, ASNS</td>
<td>6.38 × 10^{-4}</td>
<td>eNOS Signalling</td>
<td>BDKB2, PIK3R3, AQP7, PPKG1, ITPR2, PIK3R1, CHRNA9, AQP4</td>
<td>2.34 × 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPAR Signalling Pathway</td>
<td>ACOX2, AQP7, FABP3, SCD</td>
<td>7.25 × 10^{-4}</td>
<td>Claudrin-mediated Endocytosis Signalling</td>
<td>PIK3R3, ALB, CD2AP, TF, PIK3R1, TFFC, ITGB8, FGF1</td>
<td>2.95 × 10^{-5}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin Signalling Pathway</td>
<td>GYS2, PIK3R3, TRIP10, PIK3R3, PPIR3C</td>
<td>1.00 × 10^{-1}</td>
<td>ILK Signalling</td>
<td>PIK3R3, FOS, RN3, PIK3R1, ITGB6, MYL6B, MYH7B, ACTN3</td>
<td>2.95 × 10^{-5}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemostasis</td>
<td>ANGPT1, PIK3R1, ALDOA, ITPR2, PLEK, TF, ALB</td>
<td>1.64 × 10^{-3}</td>
<td>CXC4 Signalling</td>
<td>PIK3R3, FOS, RN3, ITPR2, PIK3R1, MYL6B</td>
<td>4.79 × 10^{-4}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The pathway analysis revealed a significant overrepresentation of several pathways associated with metabolic and cellular functions, such as lipid metabolism, carbohydrate metabolism, and cell signaling. The results are shown in Table 2, which summarizes the most significantly overrepresented pathways identified using Babelomics IPA and IPA. The pathways include lipid metabolism, carbohydrate metabolism, cell signaling, and metabolic disease pathways, all of which are crucial for understanding the molecular mechanisms underlying the biological effects of fatty acid composition differences.

**Discussion**

To date, muscle transcriptome analyses concerning meat quality in swine have mainly been conducted using microarrays [3,27,30–32]. Compared with microarrays, RNA-Seq enables to determine the transcript abundance with a larger dynamic range of expression levels, it is not limited by the available genomic sequencing information during microarray production and can provide additional information about new isoforms. However, the main RNA-Seq drawback when compared with microarrays is that the analysis relies on the current pig genome assembly (Scrofa10.2). Our percentage of mapped reads (85.1%) is similar to the described in the pig adipose tissue transcriptome (80%–87%) [5] performed with the Scrofa10.2 annotation version, however it was higher than the percentage found in the pig muscle transcriptome (64.8%) [6] performed with the Scrofa9.2 version or the pig liver transcriptome (71.4%–77.7%) [4] using the Scrofa9.6 annotation version. The high amount of transcripts mapping to intergenic regions and the novel coding gene discovery, showing a higher percentage of computationally predicted proteins (60.2%) versus known proteins (39.8%) in pig in comparison to other species such as bovine and human, reinforces the need to improve the current pig annotation. Similar results were shown in the porcine liver [4] and adipose tissue [5] transcriptomes, in which the 86.0% and 62.5% of the novel predicted proteins identified respectively were computationally-predicted. As expected, the major overlap of predicted novel proteins was between muscle and adipose tissue [5] because unlike the liver and gonads [2,4], both analyses were performed using the most recent annotation of the genome. Hence, a total of 40% novel predicted proteins in the muscle transcriptome have also been found in adipose tissue, either realised with the Scrofa10.2. Of the 2,372 predicted novel proteins, 972 were validated in silico being present in at least one of the three tissues compared [2,4,5]. Interestingly, 36 of the novel predicted proteins were also identified in four different tissues (liver [4], gonads [2], adipose [5], and muscle tissue) (Figure 4). When analysing the main metabolic pathways for the novel transcripts identified, the “phosphatidylinositol signalling system” and “inositol phosphate metabolism” were among the most represented categories. The phosphatidylinositol signalling system plays a critical role in the regulation of diverse processes such as muscle contraction, cell secretion, cell growth and differentiation [33]. Moreover, phosphatidylinositol is an essential component of the lipid membrane, where the total amount of phospholipids remains fairly constant, or increases little, as the animal increases in fatness [9,34]. Not surprisingly, these pathways were also identified when analyzing the adipose tissue novel transcripts [5]. Interestingly, the phosphatidylinositol signalling was also found within the most significantly overrepresented pathways in animals differing in FA composition in an study using microarrays [33]. Finally, we detected a high percentage (34.5%) of new repetitive elements present in the porcine genome. This provides important insights into the non-coding genome and its potential roles in gene regulation and function. The results also suggest the need for further studies to develop more robust and accurate annotation methods for the pig genome, which would enable a more comprehensive understanding of the molecular basis of meat quality traits.

**Muscle transcriptome description**

In the present study, using RNA-Seq analysis we were able to map a high percentage of reads to the current pig genome assembly (Scrofa10.2). Our percentage of mapped reads (85.1%) is similar to the described in the pig adipose tissue transcriptome (80%–87%) [5] performed with the Scrofa10.2 annotation version, however it was higher than the percentage found in the pig muscle transcriptome (64.8%) [6] performed with the Scrofa9.2 version or the pig liver transcriptome (71.4%–77.7%) [4] using the Scrofa9.6 annotation version. The high amount of transcripts mapping to intergenic regions and the novel coding gene discovery, showing a higher percentage of computationally predicted proteins (60.2%) versus known proteins (39.8%) in pig in comparison to other species such as bovine and human, reinforces the need to improve the current pig annotation. Similar results were shown in the porcine liver [4] and adipose tissue [5] transcriptomes, in which the 86.0% and 62.5% of the novel predicted proteins identified respectively were computationally-predicted. As expected, the major overlap of predicted novel proteins was between muscle and adipose tissue [5] because unlike the liver and gonads [2,4], both analyses were performed using the most recent annotation of the genome. Hence, a total of 40% novel predicted proteins in the muscle transcriptome have also been found in adipose tissue, either realised with the Scrofa10.2. Of the 2,372 predicted novel proteins, 972 were validated in silico being present in at least one of the three tissues compared [2,4,5]. Interestingly, 36 of the novel predicted proteins were also identified in four different tissues (liver [4], gonads [2], adipose [5], and muscle tissue) (Figure 4). When analysing the main metabolic pathways for the novel transcripts identified, the “phosphatidylinositol signalling system” and “inositol phosphate metabolism” were among the most represented categories. The phosphatidylinositol signalling system plays a critical role in the regulation of diverse processes such as muscle contraction, cell secretion, cell growth and differentiation [33]. Moreover, phosphatidylinositol is an essential component of the lipid membrane, where the total amount of phospholipids remains fairly constant, or increases little, as the animal increases in fatness [9,34]. Not surprisingly, these pathways were also identified when analyzing the adipose tissue novel transcripts [5]. Interestingly, the phosphatidylinositol signalling was also found within the most significantly overrepresented pathways in animals differing in FA composition in an study using microarrays [33]. Finally, we detected a high percentage (34.5%) of new repetitive elements present in the porcine genome. This provides important insights into the non-coding genome and its potential roles in gene regulation and function. The results also suggest the need for further studies to develop more robust and accurate annotation methods for the pig genome, which would enable a more comprehensive understanding of the molecular basis of meat quality traits.
that used an older version of the pig genome assembly [2,4,5]. This higher content of repetitive elements can be explained by the improvement of the current assembly (Sscrofa10.2) of the pig genome being the repetitive regions the most difficult to assemble [36].

Differential Expression analysis

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

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

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

Among the top canonical pathways overrepresented between both groups of animals, we found hemostasis, nitric oxide (NO), metabolism of lipids and lipoproteins and PPAR and insulin signalling pathways (Table 2). Remarkably, most of the genes represented in these pathways were down-expressed in the H group. When compared with a previous study using microarrays of animals extreme for FA metabolism [27,35] such as ACACA, FABP4 or FABP5, we used a strict single list in order to avoid false positives [39]. For the same reason we used a strict $P$-value $\leq 0.01$, based on the Q-Q plot and equivalent to a FDR $\leq 0.12$, and a fold change $\geq 1.2$ as used in the adipose transcriptome analysis [3,40]. We identified a lower number of differentially-expressed genes with DESeq (208 genes) when compared to EdgeR (314 genes) what is in accordance with the observations reported by Soneson & Delorenzi [40]. DESeq uses conservative default settings and performs well when outliers are introduced, having a better false discovery rate (FDR) control for large sample sizes than EdgeR [40]. Notice that some relevant genes identified using microarrays in the muscle transcriptome of animals extreme for FA metabolism [27,35] such as ACACA, FABP4 or FABP5 were found incorrectly or non-annotated in the Sscrofa10.2 annotation version. On the other hand, interesting genes detected in our RNA-Seq study that may determine differences in FA composition in muscle such as ChREBP, GYS2, PLIN1, PLIN5 and AQP4 could not be detected in microarray studies since probes for these genes were not included.

Differentially modulated metabolic pathways between groups

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

- **NO and insulin signalling pathways.** The skeletal muscle is a target organ for the insulin-induced glucose uptake and for the maintenance of glucose homeostasis in blood [42]. Insulin acts in the carbohydrate metabolism facilitating the glucose diffusion into adipose and muscle cells via glucose transporter proteins (GLUT) and stimulates FA synthesis and the storage of triglycerides by the esterification of glycerol phosphate. Notably, the C16:1 n-7 FA, observed to be decreased in the H group (Table 1), can act as a lipokine that jointly with the expression of the *peroxisome proliferator-activated receptor gamma* (PPAR-$\gamma$) can strongly stimulate the muscle insulin action [43,44]. Interestingly, PPAR-$\gamma$ and solute carrier family 2, member 4 (SLC2A4; also called GLUT4) were identified as over-expressed in the L group animals when using EdgeR program. Supporting these results, Canovas et al. [33] identified a higher expression of myocyte enhancer factor 2A (MEF2A) gene which upregulates GLUT4 in Duroc animals having a higher MUFA and SFA content. Furthermore, insulin stimulates eNOS, the enzyme responsible for synthesizing NO by calcium-independent phosphorylation via phosphoinositide 3-kinases (PI3 kinases) and the downstream effector serine/threonine kinase (Akt) [45]. NO is a signalling molecule synthesized from L-arginine that plays an important role in regulating energy metabolism in mammals [46]. It has been reported that a chronic exposure of NO may decrease whole-body energy metabolism, increasing the adiposity and obesity [46]. For instance, PI3 kinases, down-expressed in the H group, have been reported to be necessary for the insulin-stimulated glucose uptake and glycogen synthesis, meanwhile, Akt regulates cell growth and metabolism and it is involved in glucose transport and lipogenesis (Figure 4) [16,47]. In the same direction, the glycogen synthase (GYS) gene was down-expressed in animals belonging to the H group (Figure S1), what might be a downstream effect of the Akt pathway [48]. Thus, the GYS inhibition may decrease the synthesis of glycogen necessary for glucose storage.

- **PPAR and metabolism of lipids and lipoproteins pathways.** In concordance with the low glucose-uptake that seems to occur in the H group of animals, we observed a down-expression of lipogenic genes most probably due to the lack of activation of carbohydrate responsive-element binding protein (ChREBP) [49]. The *stearyl-CoA desaturase (SCD)* gene (Table S5) is responsible for the biosynthesis of MUFA from SFA, and its deficiency has been associated with lean mice [50]. Furthermore, polymorphisms in SCD gene have been strongly associated with FA composition in pigs and cows [12,51–53]. It has been suggested that an inhibition of this enzyme produces an increase in fatty acid oxidation through the inhibition of acetyl-CoA carboxylase (ACACA), regulated via

![Figure 4. Venn diagram of the novel predicted proteins expressed in muscle, liver [4], gonads [2] and adipose tissue [5].](image)
Another gene whose disruption is associated with lean mice and was also down-expressed in H group is the perilipin [55]. Perilipins modulate the hydrolysis of triglycerides by hormone-sensitive lipase (LIP) [56]. Specifically, perilipin 5 (PLIN5) may play a role of “master lipolytic regulator” in muscle, and its over-expression can increase lipid droplet size and triacylglycerol storage [57]. We also identified the lipid transporter fatty acid binding protein 3, muscle and heart (FABP3) and the Aquaporins (AQP4 and AQP7) as down-expressed in the H group (Table S4). FABP3 is more expressed in the skeletal muscle than in other tissues and participates in FA uptake and cytosolic transport, having a high binding affinity for palmitic, oleic and stearic acids. Furthermore, FABP3 acts as a transcription factor in the nucleus for the control of lipid-mediated transcriptional programs via nuclear hormone receptors or other transcription factors that respond to lipids [58]. This gene has also been found in a genomic region significantly associated with FA composition in a GWAS performed in the Iberian x Landrace cross, being a clear candidate to explain the differences in FA composition observed between the two groups of animals [16]. Besides, it has been suggested as a candidate gene for the control of IMF deposition as it was identified as over-expressed in animals with higher IMF content [59]. The Aquaporins are modulated by the PI3K/Akt signalling and in muscle, and its involvement is crucial for energy supply in the skeletal muscle [46,60,61]. The AQP7 which is higher expressed in fat tissue was also identified in the liver transcriptome study as being also down-expressed in animals with a higher content of PUFA [4,61]. Another differentially-expressed gene between the two groups of animals was the very long-chain specific acyl-CoA dehydrogenase gene (ACADVL), a PPAR target gene which was down-expressed in the H group. This gene catalyzes the first step of the mitochondrial FA β-oxidation pathway, mainly in muscle, having preference for C16:0, C16:1, C18:0 and C18:1 [62,63]. Moreover, ACADVL deficiency in humans produced a defective oxidation of oleic FA and knock-out mice for ACADVL fed in high-fat diet had a decrease in whole body fat content [64]. Overall, these results agree with a previous study in which the transcriptome of two groups of Duroc pigs with different IMF composition was analysed using microarrays and concluded that the IMF accumulation in animals having more IMF, MUFA and SFA may result from a balance between uptake, synthesis and degradation of triglycerides [35].

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

**Pig lipid metabolism affected by intramuscular FA composition**

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

**Implications**

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

Methods

Animal samples and phenotypes

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

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

RNA isolation

Total RNA was isolated from the Longissimus dorsi muscle of 12 samples with RiboPure Isolation of High Quality Total RNA (Ambion, Austin, TX, USA). Total RNA was quantified in a NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE, USA) and Qubit (Invitrogen, Carlsbad, CA, USA). RNA purity and integrity was checked employing a Bioanalyzer-2100 (Agilent Technologies, Inc., Santa Clara, CA, USA). All samples had a RNA Integrity Number (RIN) above 6.5.

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

Mapping and annotation

We ran FastQC [76] for the quality control. Indexed reads were then mapped to the reference pig genome version 10.2 (Sus scrofa10.2) and the annotation database Ensembl Genes 67 [77] using TopHat v2.0.1 [19] with an allowance of two mismatches for each read. The resulting bam files containing the aligned sequences, were subsequently merged with Samtools [78]. Reads were annotated using the intersectBed option of BEDTools [79]. Cufflinks v2.0.2 program [20] was used to assemble the transcripts with a minimum of 10 reads per alignment. Finally, Samtools [70] was employed to compute descriptive statistics.

Gene expression quantification and differential-expression analysis

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

Transposable elements and orthology analysis

We used RepeatMasker version open-3.3.0 [21] with the rmr20120418 database in order to identify repetitive and transposable elements in the pig muscle transcriptome. We used “quick search” and “pig” species options and the Search Engine NCBI/RMBLAST.

Intergenic expressed regions not annotated in the Sus scrofa10.2 version assembly were identified using Cuffcompare [82] and extracted using our own Python and R scripts. Novel putative proteins were predicted with Augustus program [22]. Afterwards, using Blast2GO [23], we mapped and annotated the novel protein coding genes. Using BLASTP option (E-value hit filter 1.00E-6, annotation cutoff 55, gene ontology (GO) weight 5 and HSP-hit coverage cutoff 0) we checked their orthology with already annotated proteins in Homo sapiens, Bos taurus and Sus scrofa protein databases. The InterProScan specific tool implemented in Blast2GO was employed to refine the functional annotations. With the GO Slim options we selected the relevant GO terms belonging to the cellular component, biological process and molecular function categories. Parameters were set to 10 for the seq filter and 20 for node score filter. Finally the ontology level was set to 3.

Gene ontologies and pathways

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

Data Availability

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

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

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

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

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

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

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

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

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

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

Acknowledgments

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

Author Contributions

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

References


