RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles

T. F. Cardoso1,2, A. Cánovas3, O. Canela-Xandri3, R. González-Prendes2, M. Amills1,4 & R. Quintanilla3

We have used a RNA-seq approach to investigate differential expression in the skeletal muscle of swine (N = 52) with divergent lipid profiles i.e. HIGH (increased intramuscular fat and muscle saturated and monounsaturated fatty acid contents, higher serum lipid concentrations and fatness) and LOW pigs (leaner and with an increased muscle polyunsaturated fatty acid content). The number of mRNAs and non-coding RNAs (ncRNAs) expressed in the porcine gluteus medius muscle were 18,104 and 1,558, respectively. At the nominal level of significance (P-value ≤ 0.05), we detected 1,430 mRNA and 12 non-coding RNA (ncRNA) transcripts as differentially expressed (DE) in the gluteus medius muscle of HIGH vs LOW pigs. This smaller contribution of ncRNAs to differential expression may have biological and technical reasons. We performed a second analysis, that was more stringent (P-value ≤ 0.01 and fold-change ≥ 1.5), and only 96 and 0 mRNA-and ncRNA-encoding genes happened to be DE, respectively. The subset of DE mRNA genes was enriched in pathways related with lipid (lipogenesis and triacylglycerol degradation) and glucose metabolism. Moreover, HIGH pigs showed a more lipogenic profile than their LOW counterparts.

Several RNA-seq studies have been carried out on different pig breeds in order to identify genes involved in fat deposition and meat quality1,2. Besides analysing gene expression differences, these studies aimed to dissect the complex networks of pathways and genes that determine porcine phenotypes of economic interest. In this way, the expression patterns of porcine liver, longissimus dorsi and abdominal fat were examined in two full-sib hybrid pigs with extreme phenotypes for growth and fatness traits3. The proportion of tissue-specific mRNA transcripts happened to be quite modest (<10%) and several microRNAs (miRNAs) were differentially expressed (DE) across tissues. Other studies analysing differential gene expression in muscle, fat and liver tissues of Iberian x Landrace pigs with extreme phenotypes for muscle fatty acid (FA) composition revealed that DE loci are integrated in common pathways related with LXR/RXR activation, peroxisome proliferator-activated receptors (PPARs) and β-oxidation1,4,5. A recent analysis comparing Iberian and Iberian x Duroc pigs also identified LXR/RXR activation and cholesterol synthesis as enriched pathways in the set of DE genes2. In contrast, the potential role of ncRNAs in muscle fat deposition has been scarcely studied in pigs4,6.

In a previous experiment, we demonstrated that genes involved in FA uptake, lipogenesis, triacylglycerol synthesis, lipolysis and insulin signalling are DE in the skeletal muscle of Duroc pigs with divergent lipid phenotypes7. One drawback of this study was that gene expression was measured with microarrays, which have a limited dynamic range, sensitivity (specially for low-abundance transcripts) and specificity. Moreover, the expression of non-coding RNAs could not be measured with Affymetrix porcine microarrays. In the current work, we aimed to circumvent all these limitations by analysing, through a RNA-seq approach, the muscle transcriptome of a subset of Duroc pigs with distinct lipid profiles.
of these Duroc pigs. Our goal was to determine the relative contributions of protein-coding and non-coding RNAs to differential expression in the skeletal muscle of pigs with distinct lipid profiles.

Results
The RNA-seq experiment allowed us generating an average of 133 million paired-end reads per sample and 72.8% of them were successfully mapped to the pig S.s.crofa10.2 genome assembly. The percentages of exonic and intronic reads were 91.4% and 8.6%, respectively. After quality control analysis, four samples were discarded. Thereby, we used a final dataset of 26 animals per group (HIGH and LOW) to identify DE genes.

Differential expression of mRNA encoding genes. A total of 1,430 mRNA genes happened to be DE when considering exclusively a significance threshold of $P$-value $\leq 0.05$ (Supplementary Table S1). Only 76 of these 1,430 mRNA-encoding genes were identified as DE by Cánovas et al. when they compared the gene-expression of HIGH and LOW pigs retrieved from the same population employed by us (Supplementary Figure 1, Supplementary Table S2). When we performed a more stringent analysis ($q$-value $\leq 0.05$ and fold-change $\leq 1.5$), 76 genes were DE (Supplementary Table S3). Moreover, twenty-one genes remained significant after correction for multiple testing ($q$-value $\leq 0.05$ and fold-change $\geq 1.5$), as shown in Table 1.

We used the IPA package (QIAGEN Redwood City, www.qiagen.com/ingenuity) to identify pathways to which DE genes belong to as well as to explore the existence of signalling networks connecting DE genes. Forty four pathways were significantly enriched in the dataset of 96 DE genes (Supplementary Table S4). This information should be interpreted with caution because, in general, pathways were represented by a small number of genes and statistical significance was not very high. Amongst the enriched pathways, it is worth to mention TR/RXR activation, synthesis of palmitate and stearate, FA biosynthesis, triglycerol degradation, and the conversion of acetate into acetyl-CoA (Table 2, Supplementary Table S4). A complementary analysis with the ReactomeFIViz app revealed 50 significant pathways (Supplementary Table S5). Differentially expressed mRNA genes were also grouped in gene regulatory networks with the IPA software. As shown in Supplementary Table S6, we found eleven regulatory networks related with a variety of functions, and the top-scoring one was that of Cardiovascular Disease, Cardiovascular System Development and Function, Organismal Injury and Abnormalities (Fig. 1 and Supplementary Table S6).

The Regulator Effects tool of the IPA package was employed to identify potential transcriptional regulators that may explain the differential patterns of expression observed between HIGH and LOW pigs (Fig. 2). By doing so, two main transcriptional regulators were identified i.e. peroxisome proliferator-activated receptor γ (PPARG) and platelet-derived growth factor BB (PDGFB). In the network shown in Fig. 2, these genes appear to be involved in an heterogeneous array of biological functions related with the quantity of carbohydrate, insulin sensitivity, necrosis of prostate cancer cell lines and apoptosis of lymphocytes. Indeed, the PPARG gene ($P$-value = 0.02 and FC = 1.36) is depicted as a key regulator of genes related with carbohydrate metabolism (CEBPA, CES1, CIDEC) and the inhibition of insulin sensitivity (CES1, CIDEC, FASN).

| Ensembl ID                | Gene name | Fold-Change | $P$-value | $q$-value |
|---------------------------|-----------|-------------|-----------|-----------|
| ENSSSCG00000005648        | SLC27A4   | 1.66        | 1.32E-06  | 4.28E-03  |
| ENSSSCG00000027946        | MYP       | 1.78        | 2.63E-06  | 5.97E-03  |
| ENSSSCG00000017232        | SLC9A3R1  | 1.72        | 1.26E-05  | 1.36E-02  |
| ENSSSCG00000005935        | AGO2      | 1.59        | 1.77E-05  | 1.43E-02  |
| ENSSSCG0000003379         | KLHL21    | 1.79        | 1.61E-05  | 1.43E-02  |
| ENSSSCG00000011740        | SERPIN1   | $-1.81$     | 2.48E-05  | 1.72E-02  |
| ENSSSCG0000001931         | GRAMD2    | $-1.58$     | 2.74E-05  | 1.76E-02  |
| ENSSSCG0000001444         | NT5DC2    | 1.54        | 3.26E-05  | 1.76E-02  |
| ENSSSCG0000007574         | SDK1      | 1.58        | 2.97E-05  | 1.76E-02  |
| ENSSSCG0000007745         | SUMF2     | $-1.54$     | 4.21E-05  | 1.95E-02  |
| ENSSSCG0000002939         | ITGA5     | 1.72        | 4.46E-05  | 1.96E-02  |
| ENSSSCG0000007133         | ACS51     | 1.51        | 5.18E-05  | 2.09E-02  |
| ENSSSCG0000288814         | SOD3      | 1.97        | 5.37E-05  | 2.09E-02  |
| ENSSSCG0000006277         | SPIDR     | 2.04        | 5.90E-05  | 2.19E-02  |
| ENSSSCG0000007554         | ZFAND2A   | 2.54        | 1.01E-04  | 2.88E-02  |
| ENSSSCG0000003105         | SLCL5A    | 1.67        | 1.17E-04  | 3.07E-02  |
| ENSSSCG0000010529         | SFRP5     | 2.03        | 1.31E-04  | 3.11E-02  |
| ENSSSCG0000006245         | SDR16C5   | 3.02        | 1.36E-04  | 3.15E-02  |
| ENSSSCG0000013579         | CD209     | 1.95        | 1.50E-04  | 3.31E-02  |
| ENSSSCG0000008232         | RNF181    | $-2.09$     | 2.05E-04  | 3.57E-02  |
| ENSSSCG00000030165        | MAFF      | 1.67        | 2.22E-04  | 3.72E-02  |

Table 1. List of the most significant differentially expressed genes in HIGH and LOW pigs after correcting for multiple testing ($q$-value $\leq 0.05$ and fold-change $\geq 1.5$). A negative FC means that the affected gene is overexpressed in LOW pigs.
**Differential expression of non-coding RNAs.** We identified 1,558 ncRNA transcripts expressed in the pig *gluteus medius* muscle, with sizes between 53 and 9,032 bp (Supplementary Table S7). Amongst these, 1,354 and 204 transcripts were classified as small (snRNA) and long (lnRNA) non-coding RNAs, respectively. It is important to emphasize that the annotation of porcine ncRNAs is still very preliminary and it should be taken with caution. In general, snRNA had orthologous sequences in other mammalian species, while lnRNAs were much less conserved (Table 3). We only detected 12 ncRNAs (11 lnRNAs and 1 snRNA) that were DE at the nominal level ($P$-value $\leq 0.05$), while none of these ncRNAs remained significant after correction for multiple testing (in all cases the $q$-value was non-significant, Table 4).

In addition, we identified 25 mRNA-encoding genes that mapped near (30 kb or less) to the subset of DE ncRNA loci (Table 5). This observation may have biological implications because ncRNAs often cis-regulate the expression of genes located in their vicinity. Within this list of neighbouring genes (Table 5), CU468594.8 ($P$-value $=0.003$ and FC $=1.26$) and MT-ND6 ($P$-value $=0.038$ and FC $=−1.21$) mRNAs are DE in HIGH vs LOW pigs ($P$-value $<0.05$ and 1.2-fold change in expression).

**Discussion**

**Divergent muscle mRNA expression profiles in pigs with extreme phenotypes for fatness traits.** After correcting for multiple testing, twenty-one genes, displaying a wide array of functional roles, showed a significant DE between HIGH and LOW pigs (Table 1). For instance, *SLC27A4* is involved in the translocation of long-chain fatty acids across the plasma membrane while *SERP5* plays a role in anti-inflammatory and insulin-sensitizing processes and *MVP* and *AGO2* contribute to RNA interference and signal transduction and transport, respectively. Two of the genes listed in Table 1 might be related with meat quality i.e. *RNF181*, which encodes a E3 ubiquitin-protein ligase that participates in the degradation of muscle proteins through the ubiquitin-proteasome system, and *SDK1*, which has been associated with intramuscular fat (IMF) content in Large White pigs.

The Spearman correlation between the microarray data reported by Canovas et al. in 68 HIGH and LOW pigs and RNA-seq data generated in the current study (N = 52) was 0.54. This value is comparable to what has been published in previous studies analysing gene expression in human brain cells ($r = 0.61–0.67$) and proliferating vs quiescent fibroblasts ($r = 0.18–0.42$). We also compared our dataset of DE genes with those detected by Canovas et al. As shown in Suppl Figure 1 the level of concordance was quite low (only 76 genes were simultaneously identified by both platforms). A modest overlap between microarray and RNA-seq data has been reported in previous studies. For instance, Trost et al. analysed the concordance between both types of data in fibroblasts cultured at two different developmental stages, and they just found an overlap of around 25% in the two lists of DE genes. This value is higher than the one reported by us, but it is important to highlight that the analysis of Trost et al. was based on a set of probes common to both platforms. Moreover, the microarray analysis performed by Canovas et al. was based on a dataset of around 68 pigs, while we used a subset of 52 individuals in our RNA-seq analysis. Trost et al. used quantitative real-time PCR as a third approach to validate microarray and RNA-seq data and they found that RNA-seq outperforms the microarray technology. However, differences between both methods are not dramatic i.e. the Spearman correlations between microarray and RNA-seq data vs qPCR validation results were 0.44 and 0.56, respectively. This means that both technologies detect different sets of DE expressed genes and, in consequence, they are complementary. According to Wang et al., the magnitude of the treatment effect has a strong impact on the level of concordance between microarray and RNA-seq platforms i.e. large discrepancies can be anticipated when two similar biological conditions are compared. Low-abundance transcripts are another source of discrepancy between both methodological approaches.

We found some evidence that pathways related with lipid synthesis (stearate, palmitate and FA synthesis) and catabolism (triacylglycerol degradation), glucose metabolism (glucose synthesis and degradation) and hormonal response (growth hormone signalling) were enriched in the set of DE genes (Table 2 and Supplementary Table S4). Similar results were obtained by Cánovas et al. i.e. they detected an overexpression of pathways related with the synthesis of FA and insulin signaling in HIGH pigs. Puig-Oliveras et al. compared the muscle mRNA expression of pigs with high saturated (SFA) and monounsaturated (MUFA) FA muscle contents against those with a high polyunsaturated FA (PUFA) content and also observed an enrichment of pathways related with fat deposition.
(PPAR and insulin signalling) in the set of DE genes. Insulin stimulates the absorption of glucose, which is a lipogenic substrate, and PPAR enhances triglyceride storage\(^{19}\). By using the same animal material employed by Puig-Oliveras et al.\(^1\), Corominas et al.\(^5\) observed an overexpression of genes belonging to the LXR/RXR activation pathway in the adipose tissue of pigs with high muscle SFA and MUFA contents. These results, which agree well with ours (Supplementary Table S4), make sense because liver X receptors are sterol-activated transcription factors that enhance lipogenesis\(^{20}\).

Though not all studies comparing pigs with divergent lipid phenotypes identify the same sets of pathways, an outcome that partly depends on the software and databases used as well as on the targeted tissue and phenotype variability, the general trend that emerges is that biochemical routes that promote lipid deposition are overexpressed in the skeletal muscle of fat pigs with high muscle SFA and MUFA contents. In close concordance with a previous study\(^7\), we have also found that one gene that promotes the catabolism of triglycerides, carboxylesterase 1 (\(CES1\)), is strongly upregulated in HIGH pigs (\(P\)-value = 0.0006, FC = 2.4). The CES1 protein has hydrolase activity and its inactivation leads to hyperlipidemia and increased fat deposition in peripheral tissues, obesity, fatty liver, hyperinsulinemia and insulin insensitivity and a decreased energy expenditure\(^{21}\). According to Cánovas and coworkers\(^7\), the upregulation of lipolytic genes in HIGH pigs suggests the existence of a cycle where triacylglycerols are continuously synthesized and degraded. However, we have also detected the downregulation of lipolytic genes such as lipase C, hepatic type (\(LIPC\), \(P\)-value = 0.002, FC = −1.5)\(^{22}\), a feature that suggests that the mechanisms that promote an adequate balance between anabolic and catabolic lipid metabolism routes are highly complex.

Analysis of the data with the IPA software (QIAGEN) showed that the top-scoring regulatory network was Cardiovascular Disease, Cardiovascular System Development and Function, Organismal Injury and Abnormalities, a result that is not surprising given the tight relationship between lipoprotein metabolism and cardiovascular risk\(^{23}\).

In the network shown in Fig. 1, the V-Akt murine thymoma viral oncogene homolog molecule (AKT) occupies a central position, having connections with several DE lipid-related genes (e.g., \(TRIB3\), \(TIMP1\) and \(ITGA5\)). Interestingly, AKT is one of the main regulators of glucose homeostasis\(^{24}\), a feature that suggests that the mechanisms that promote an adequate balance between anabolic and catabolic lipid metabolism routes are highly complex.

When we used the Regulator Effects tool of IPA, the \(PPARG\) and \(PDGFB\) genes were predicted to be major transcriptional regulators of the set of 96 DE loci (Fig. 2). The \(PPARG\) transcription factor is critically required for adipogenesis, being a powerful modulator of whole-body lipid homeostasis and insulin sensitivity\(^{25}\). Polymorphism in the \(PPARG\) gene is associated with individual susceptibility to type 2 diabetes, obesity and body mass index\(^{26}\). In our study, \(PPARG\) is upregulated (\(P\)-value = 0.02 and FC = 1.36) in HIGH pigs and appears to regulate several genes, such as \(CEBPA\) (\(P\)-value = 0.009 and FC = 1.64), \(CES1\) (\(P\)-value = 0.0004 and FC = 2.03),...
CIDEC (P-value = 0.0005 and FC = 2.46) and FASN (P-value = 0.0009 and FC = 2), that play distinct roles in lipid metabolism (http://www.genome.jp/kegg/pathway.html).

**Limited contribution of the non-coding RNA transcriptome to differential expression between HIGH and LOW pigs.** Non-coding RNAs have been shown to regulate gene expression by interacting with chromatin complexes, working as RNA enhancers, recruiting or assembling certain proteins and interacting with other RNAs at the post-transcriptional level\textsuperscript{27}. In consequence they may play a fundamental role in the metabolism of the porcine skeletal muscle. In our study, we have identified 1,558 muscle-expressed ncRNA transcripts

| Transcript | Transcript Type | Number | Conserved ncRNA |
|------------|----------------|--------|-----------------|
| Small ncRNA | miRNA | 433 | 137 |
| | miscRNA | 95 | 82 |
| | Mt-rRNA | 2 | 0 |
| | Mt-tRNA | 22 | 0 |
| | rRNA | 57 | 52 |
| | snoRNA | 417 | 395 |
| | snRNA | 328 | 273 |
| Long ncRNA | Non coding | 4 | 0 |
| | Processed transcript | 143 | 0 |
| | Antisense | 15 | 0 |
| | lincRNA | 42 | 0 |

Table 3. **Evolutionary conservation of non-coding RNAs transcribed in the porcine gluteus medius muscle.** miRNA = microRNAs; misc_RNA = miscellaneous other RNA; Mt-rRNA = Mitochondrial ribosomal RNA; Mt-tRNA = transfer RNA located in the mitochondrial genome; rRNA = ribosomal RNA; snoRNA = small nucleolar RNA; snRNA = small nuclear RNA; lincRNA = Long intergenic non-coding RNAs.
The total number of ncRNAs in the pig genome is currently unknown, but Zhou et al.\textsuperscript{28} highlighted the existence of at least 6,621 long intergenic non-coding RNAs (lincRNA) transcripts encoded by 4,515 gene loci. In humans, 58,648 lincRNA encoding loci have been identified so far\textsuperscript{29}. In our dataset (Table 3), the degree of evolutionary conservation of sncRNAs happened to be much higher than that of lncRNAs. Zhou et al.\textsuperscript{28} characterized the porcine lincRNA transcriptome and found that only 40% of the transcripts had a detectable human lincRNA ortholog. This scarcity of orthologous sequences can be due, in part, to the poor annotation of ncRNAs in all investigated species.

(Supplementary Table S7).
There is growing evidence that there might be a positive correlation between the expression of ncRNAs and nearby mRNA encoding genes, suggesting that the former may regulate the expression of the latter⁶. We investigated this issue by analysing if there are DE protein-coding genes in the vicinity of any of the 12 DE ncRNAs identified in our work (P-value ≤ 0.05, Tables 4 and 5). Two protein-coding genes, i.e. mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6 (MT-ND6) and CUA68594.8, fulfilled this condition (P-value ≤ 0.05 and FC ≥ 1.2, Table 5). The MT-ND6 gene encodes a NADH dehydrogenase that catalyses the oxidation of NADH by ubiquinone, an essential step in the mitochondrial electron transport chain¹⁴. The CUA68594.8 locus is orthologous to human solute carrier family 52-riboflavin transporter, member 2 (SLC52A2). Riboflavin is the precursor of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), two essential cofactors that participate in a wide range of redox reactions²²,²³.

We aimed to ascertain if differences amongst HIGH and LOW pigs, in terms of IMF content and composition, are mainly due to the DE of either mRNA or ncRNA encoding genes. When considering a nominal P-value of 0.05 as a threshold of significance, the number of DE ncRNAs (12 loci) was much smaller than that of DE mRNAs (1,430 loci), even if we take into account that the number of expressed mRNAs (18,104) was also higher than that of ncRNAs (1,558). Moreover, none of the DE ncRNAs remained significant after correction for multiple testing. In a recent experiment, the transcriptome of pig endometrial samples collected at different pregnancy stages was characterized, and 2,376 transcripts were identified as DE in pairwise comparisons⁴⁴. Only 12% of these transcripts corresponded to IncRNAs indicating that changes in the endometrial transcriptome associated with pregnancy mainly affect the expression of protein-coding genes. However, studies performed in humans indicate a much more balanced contribution of mRNAs and ncRNAs to differential expression. For instance, Wang et al.⁵³ investigated the expression patterns of peripheral leukocytes of healthy and autistic individuals and identified 3,929 and 2,591 DE lncRNAs and mRNAs, respectively. Similarly, Zhou et al.⁵⁵ identified 891 and 576 DE mRNAs and IncRNAs, respectively, when comparing the expression patterns of ectopic and eutopic endometrial tissue. These differences between humans and pigs are probably the consequence of technical rather than biological causes, evidencing the pressing need of improving the genomic and functional annotation of porcine ncRNAs.

Conclusions

By comparing the mRNA expression of HIGH and LOW pigs by RNA-seq, we have identified 96 loci displaying differential expression (P-value ≤ 0.01 and FC ≥ 1.5). Many of these loci were not detected in a previous microarray-based experiment, suggesting that distinct platforms detect different sets of DE genes. Lipid biosynthetic pathways were enriched in DE genes and upregulated in HIGH pigs, a result that is consistent with previous reports. We have also undertaken the analysis of non-coding RNAs, a feature that has been neglected in previous studies investigating the differential expression of porcine genes. Our results indicate that the number of DE non-coding RNAs is much lower than that of mRNAs, an outcome that might be partly explained by the poor annotation of porcine ncRNAs.

Material and Methods

Ethics statement. All experiments were performed in accordance with the ARRIVE guidelines (https://www.nc3rs.org.uk/arrive-guidelines). Animal care and management procedures were approved by the Ethical Committee of the Institut de Recerca i Tecnologia Agroalimentaries, IRTA.

Animal Material. One population of 350 Duroc barrows belonging to 5 half-sib families, and distributed in 4 fattening batches was generated in 2003. All animals were kept under the same feeding and management conditions³⁷. A wide array of growth, fatness, feed efficiency and carcass and meat quality traits were recorded in these animals, including weight, daily food intake, fat deposition, and IMF content and composition (C:12-C:22 interval) of the gluteus medius muscle.³⁸ Many of these loci were not detected in a previous microarray-based experiment, suggesting that distinct platforms detect different sets of DE genes. Lipid biosynthetic pathways were enriched in DE genes and upregulated in HIGH pigs, a result that is consistent with previous reports. We have also undertaken the analysis of non-coding RNAs, a feature that has been neglected in previous studies investigating the differential expression of porcine genes. Our results indicate that the number of DE non-coding RNAs is much lower than that of mRNAs, an outcome that might be partly explained by the poor annotation of porcine ncRNAs.

RNA isolation and library construction and sequencing. Total RNA was isolated from 56 porcine gluteus medius muscle samples (28 HIGH and 28 LOW) by using the acid phenol method implemented in the RiboPure kit (Ambion, Austin, TX). Total RNA was quantified in a Nanodrop ND-1000 spectrophotometer, checked for purity and integrity in a Bioanalyzer-2100 device (Agilent Technologies, Inc., Santa Clara, CA) and checked for purity and integrity in a Bioanalyzer-2100 device (Agilent Technologies, Inc., Santa Clara, CA) and submitted to the Centre Nacional d’Anàlisi Genòmica (CNAG, http://www.cnag.cat) for sequencing. Libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina Inc) according to the protocols recommended by the manufacturer. Each library was paired-end sequenced (2 × 75 bp) by using the TruSeq SBS Kit v3-HS, in a HiSeq2000 platform.

Bioinformatic analyses. All bioinformatic analyses were performed with the CLC Bio Workbench software (CLC Bio, Aarhus, Denmark). Quality control was carried out with the NGS Core Tools, considering several parameters based on the FastQC-project (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). We carried out per-sequence and per-base analyses to filter reads according to the following criteria: sequence-read distribution = 75 bp, 100% coverage in all bases, GC-content ~50%, ~25% of A, T, G and C nucleotide contributions, ambiguous base-content < 0.1% and a Phred score higher than 30 (i.e. base-calling accuracy lower than 99.9%). Short sequence reads were assembled, mapped and annotated by using as template the pig reference genome version 10.2 (Sscrofa10.2-http://www.ensembl.org/info/data/ftp/index.html). For mapping purposes, we just considered alignments with a length fraction of 0.7 and a similarity fraction of 0.8. Besides, two mismatches and three insertions and deletions per read were allowed.
Gene expression data were normalized by calculating the reads per kilobase per million mapped reads (RPKM)\(^6\). Using scales of abundance estimates by exon length and millions of mapped reads, original expression values were transformed and normalized. More specifically, data were transformed on a decimal logarithmic scale and a scaling algorithm was utilized for the normalization of average scores\(^7\). For the statistical analysis of differential expression, we used a two-tailed t-test that assumes a Gaussian distribution and homogenous variances. This statistical test compares the mean expression levels in the two experimental groups (HIGH vs LOW) and evaluates the significance of the difference relative to the variance of the data within the groups. Multiple testing correction was performed by using a false-discovery rate approach (cut-off = 0.05) implemented in the QVALUE R package\(^8\). Fold-Change was computed as the ratio of HIGH vs LOW gene expressions (a negative FC means that the affected gene is upregulated in LOW pigs).

Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) was used to identify gene ontologies, pathways, and regulatory networks to which DE genes belong to, as well as upstream regulators. Ingenuity Pathway Analysis can transform a set of genes into a number of relevant networks based on comprehensive records maintained in the Ingenuity Pathways Knowledge Base. Networks are presented as graphs depicting the biological relationships between genes/gene products. Genes are shown as nodes, and the molecular relationship between two nodes is represented with either a solid (direct interactions) or a dashed (indirect interactions) line. The analysis of upstream regulators considers every possible transcription factor and upstream regulator contained in the Ingenuity Knowledge Base repository as well as their predicted effects on gene expression (inferred from the scientific literature). Then, this tool analyses if the patterns of expression observed in the DE genes can be explained by the activation/inhibition of any of these regulators through the calculation of a z-score i.e. a statistical measure of the match between expected relationship direction between the regulator and its targets and observed gene expression\(^9\). A parallel analysis was performed with the Cytoscape software\(^10\) by using the ReactomeFIViz app\(^9\). IPA and Cytoscape analyses were performed on a subset of DE genes, with P-value < 0.01 and a FC ≥ 1.5. Transcript classification and the search of homologs of porcine ncRNAs in other mammalian species were carried out with tools implemented in the BioMart web interface (http://www.ensembl.org/biomart/martview).

References
1. Puig-Oliveras, A. et al. Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. *PloS One* **9**, e99720 (2014).
2. 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**, e0145162 (2015).
3. Chen, C. et al. A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. *BMC Genomics* **12**, 448 (2011).
4. Ramayo-Calda, Y. et al. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics* **13**, 547 (2012).
5. Corominas, J. et al. Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. *BMC Genomics* **14**, 843 (2013).
6. Ye, R.-S. et al. Comparative anterior pituitary miRNA and mRNA expression profiles of Iama minipigs and Landrace pigs reveal potential molecular network involved in animal postnatal growth. *PloS One* **10**, e0131987 (2015).
7. Canovas, A., Quintanilla, R., Amills, M. & Pena, R. N. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics* **11**, 372 (2010).
8. Wu, G., Dawson, E., Duong, A., Haw, R. & Stein, L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. *PLoS One* **8**, 146 (2014).
9. Jia, Z., Mouison, C. L., Pie, Z., Miner, J. H. & Watkins, P. A. Fatty acid transport protein 4 is the principal very long chain fatty acyl-CoA synthetase in skin fibroblasts. *J. Biol. Chem.* **282**, 20573–83 (2007).
10. Carstensen, M. et al. Effect of Sfrp5 on cytokine release and insulin action in primary human adipocytes and skeletal muscle cells. *PloS One* **9**, e85906 (2014).
11. Kalantari, R., Chiang, C. M. & Corey, D. R. Regulation of mammalian transcription and splicing by nuclear RNAi. *Nucl. Acids Res.* **44**, 524–37 (2016).
12. Berger, W., Steiner, E., Grusch, M., Elbling, L. & Micksche, M. Vauls and the major vault protein: novel roles in signal pathway regulation and immunity. *Cell. Mol. Life Sci.* **66**, 43–61 (2009).
13. Cao, P. R., Kim, H. J. & Lecker, S. H. Ubiquitin-protein ligases in muscle wasting. *J. Anim. Breed. Genet.* **133**, 277–82 (2014).
14. Fu, X. et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**, 161 (2009).
15. Trost, B. et al. Concordance between RNA-sequencing data and DNA microarray data in transcriptome analysis of proliferative and quiescent fibroblasts. *R. Soc. Open Sci.* **2** (2015).
16. Kogenaru, S., Qing, Y., Guo, Y. & Wang, N. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics* **13**, 629 (2012).
17. Wang, C. et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* **32**, 926–32 (2014).
18. Ferrer, P. The biology of peroxisome proliferator-activated receptors: relationship with lipid metabolism and insulin sensitivity. *Diabetes* **53** Suppl 1, S43–50 (2004).
19. Hong, C. & Tontonoz, P. Liver X receptors in lipid metabolism: opportunities for drug discovery. *Nat. Rev. Drug Discov.* **13**, 433–44 (2014).
20. Quirga, A. D. et al. Deficiency of carboxylesterase 1/esterase-x results in obesity, hepatic steatosis, and hyperlipidemia. *Hepatology* **56**, 2188–98 (2012).
21. Perret, B. et al. Heparic lipase: structure/function relationship, synthesis, and regulation. *J. Lipid Res.* **43**, 1163–9 (2002).
22. Millán, J. et al. Lipoprotein ratios: Physiological significance and clinical usefulness in cardiovascular prevention. *Vasc. Health Risk Manag.* **5**, 757–65 (2009).
23. Hay, N. Akt isoforms and glucose homeostasis - the leptin connection. *Trends Endocrinol. Metab.* **22**, 66–73 (2011).
24. Ahmadian, M. et al. PPARy signaling and metabolism: the good, the bad and the future. *Nat. Med.* **99**, 557–66 (2013).
25. Masud, S. & Ye, S. Effect of the peroxisome proliferator activated receptor-γ gene Pro12Ala variant on body mass index: a meta-analysis. *J. Med. Genet.* **40**, 773–80 (2003).
RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles.

The authors declare no competing financial interests.

Acknowledgements
Part of the research presented in this publication was funded by grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of Economy and Competitiveness. We also acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Center for Research in Agricultural Genomics. The authors are indebted to Selección Batallé S.A. for providing the animal material. We gratefully acknowledge to J. Reixach (Selección Batallé) and J. Soler (IRTA) for their collaboration in the experimental protocols. Tainã F Cardoso was funded to Selección Batallé S.A. for providing the animal material. We gratefully acknowledge to J. Reixach (Selección Batallé) and J. Soler (IRTA) for their collaboration in the experimental protocols.

Author Contributions
R.Q. and M.A. conceived the study and designed the experiment; R.Q. was responsible for producing the animal material, phenotypic records and RNA-Seq data; T.F.C., A.C. and O.C. did the bioinformatic analyses; R.G.P. helped in the statistical analyses; M.A. and T.F.C. wrote the manuscript. All authors helped to draft the manuscript and read and approved its final version.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Cardoso, T.F. et al. RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. Sci. Rep. 6, 40005; doi: 10.1038/srep40005 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.