Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness

Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness

Journal:	*Animal Genetics*
Manuscript ID	AnGen-15-06-0185.R2
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Zambonelli, Paolo; Bologna University, Department of Agricultural and-Food Sciences (DISTAL) Gaffo, Enrico; University of Padova, Department of Molecular Medicine Zappaterra, Martina; Bologna University, Department of Agricultural and-Food Sciences (DISTAL) Bortoluzzi, Stefania; University of Padova, Department of Molecular Medicine Davoli, Roberta; Bologna University, Department of Agricultural and-Food Sciences (DISTAL)
Keywords:	backfat, fat deposition, gene expression, differential analysis, pigs

Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness

Zambonelli Paolo[1]*, Gaffo Enrico[2]*, Zappaterra Martina[1], Bortoluzzi Stefania[2§], Davoli Roberta[1§]

[1]Department of Agricultural and-Food Sciences (DISTAL), Bologna University, Via Fratelli Rosselli 107, 42123 Reggio Emilia, Italy.

[2]Department of Molecular Medicine, University of Padova, Via Gabelli 63, 35121 Padova, Italy

§ To whom correspondence should be addressed. Roberta Davoli Tel. +39 051 2088050, Fax +39 051 2086172, roberta.davoli@unibo.it; Stefania Bortoluzzi Tel: +39 049 8276502, Fax: +39 049 8276209, stefania.bortoluzzi@unipd.it

*These authors contributed equally to this work

Summary

Fat deposition is a widely studied trait in pigs for the implications with animal growth efficiency, technological and nutritional characteristics of meat products, but the global framework of the biological and molecular processes regulating fat deposition in pigs is still incomplete. This paper describes the backfat tissue transcription profile in Italian Large White pigs and reports genes differentially expressed between fat and lean animals according to RNA-seq data. The backfat transcription profile was characterized by the expression of 23,483 genes of which 54.1% were represented by known genes. Of 63,418 expressed transcripts, about 80% were non previously annotated isoforms. By comparing the expression level of fat *vs*. lean pigs we detected 86 robust differentially expressed transcripts, 72 more expressed (e.g. *ACP5, BCL2A1, CCR1, CD163*, *CD1A, EGR2*, *ENPP1*, *GPNMB*, *INHBB, LYZ, MSR1, OLR1, PIK3AP1, PLIN2, SPP1*, *SLC11A1, STC1*) and 14 less expressed (e.g. *ADSSL1, CDO1, DNAJB1, HSPA1A, HSPA1B, HSPA2, HSPB8, IGFBP5, OLFML3*) in fat pigs. The main functional categories enriched in differentially expressed genes were immune system process, response to stimulus, cell activation, and skeletal system development, for the overexpressed, unfolded protein binding and stress response, for the under-expressed genes, which include five heat shock proteins. Adipose tissue alterations and impaired stress response are linked to inflammation and, in turn, to adipose tissue secretory activity similarly to what is observed in human obesity. Our results open the opportunity to identify biomarkers of carcass fat traits to improve pig production chain and to identify genetic factors that regulate the observed differential expression.

Keywords: backfat, fat deposition, gene expression, differential analysis, pigs

Introduction

Backfat deposition and fat traits are among the most important characters studied in pigs, due to their strong relation with human nutrition of pig products and for the technological characteristics of high quality Protected Designation of Origin (PDO) dry-cured hams. The amount of fat laid on the external part of the pig body (subcutaneous fat or backfat) is of extreme importance for growth performances, as the lesser is the deposed fat, the better the growth performances. Regarding technological aspects related to the dry-cured high quality products and meat industry, an adequate layer of fat is required for the seasoning process of PDO products, like dry cured hams (Bosi & Russo, 2004; Čandek-Potokar & Škrlep, 2012).

During the last decade, pig transcriptomic data have been obtained initially by expressed sequence tag sequencing (Mikawa et al., 2004; Uenishi et al., 2004; Chen et al., 2006; Gorodkin et al., 2007; Uenishi et al., 2007) and microarrays (Hornshøj et al., 2007; Ferraz et al., 2008; Moon et al., 2009), which allowed the comparison of gene expression level in several pig tissues. More recently, the RNA-seq approach was used to compare the transcription profile of different pig fat tissues or different pig breeds (Chen et al., 2011; Li et al., 2012; Corominas et al., 2013; Jiang et al., 2013; Zhou et al., 2013; Sodi et al., 2014; Toedebush et al., 2014; Wang et al. 2014). The differentially expressed genes (DEG) reported in these studies are useful to investigate the metabolic pathways activated by or associated with an increased fat deposition in pig body. However, the large amount of data produced and the results reported in literature are often hardly comparable because of differences in the studied breeds; heterogeneous animals' ages; and fat deposition stages.

Moreover, these researches identified several new genes and transcripts not reported in swine or other species. To date, the global framework of the biological processes regulating backfat deposition in pigs is still incomplete, and literature is poor of studies carried out on a homogeneous sample of individuals of the same breed reared on the same environmental conditions.

The objective of this research was to investigate the transcription profile of Italian Large White (ILW) pig backfat tissue and to compare the transcriptome of animals reared in the same herd and farming conditions and showing high (FAT) and low (LEAN) backfat thickness. Moreover a first functional characterization of DEGs has been obtained to provide new insights on genes, pathways and processes influencing the divergent aptitude of subcutaneous adipose tissue deposition in ILW pigs.

Materials and methods

Samples collection and RNA extraction

We sampled twenty individuals from a purebred population of 949 ILW sib-tested pigs provided by the Italian National Association of Pig Breeders (Associazione Nazionale Allevatori Suini, ANAS, http://www.anas.it. Accessed 22 June 2015). All animals used in this study were kept according to Italian and European law for pig production and all procedures described were in compliance with national and European Union regulations for animal care and slaughtering. The animals were reared on the ANAS Sib-Test genetic station from about 30 kg live weight to at least 155 kg live weight. For the genetic evaluation of a boar, full sib triplets (two females and one castrated male) were farmed on the genetic station to be performance tested. The formula and amount of the ration was the same for all. It was

based mainly on cereals and soybean, given in excess calculated using the "*quasi ad libitum*" rule (a ration sufficiently abundant that 60% of pigs were able to ingest the full supplied food). At the end of tests, animals were transported to a commercial abattoir located about 25 km far from the test station according to the Council Rule (EC) No 1/2005 on the protection of animals during transport and related operations and amending Directives 64/432/EEC and 93/119/EC and Regulation (EC) No 1255/97. At slaughterhouse the pigs were electrical stunned and bled in a lying position in agreement with the Council Regulation (EC) No 1099/2009 on the protection of animals at the time of killing. All slaughter procedures were monitored by the Veterinary team appointed by the Italian Ministry of Health. Backfat samples were collected after slaughter, from 949 ILW pigs slaughtered at an average hot carcass weight of 118.97 kg (±0.29 SEM) and at an average age of eight months during the years 2011 and 2012 in 27 different slaughtering days. The collected samples were immediately frozen in liquid nitrogen and stored at -80°C in a deep freezer until RNA extraction. For the RNA-seq analysis we selected the animals according to the estimated breeding value (EBV) for backfat thickness (BFT) calculated by ANAS as described by Russo et al. (2000; 2008). EBVs were determined through a BLUP multiple-trait animal model procedure (Henderson & Quaas, 1976) using the BFT, measured in mm, recorded post mortem in correspondence of the *gluteus medius* muscle. The model included fixed effects of batch in test, sex, age at beginning of test, age of sow, weight at slaughter, age at slaughter, and inbreeding coefficient as well as the random effects of litter, individual permanent environment, and animal. Pigs' genetic merit for the BFT trait was calculated taking into account the additive relationship matrix. EBVs were expressed as differences from the genetic mean value for the considered trait in the year 1993. Backfat thickness genetic index may present negative values because the value of the trait is referred to the

fixed genetic base defined by ANAS as mean values of the pigs born in 1993 and considered as "zero", so the more negative values indicate lower values of BFT. The animals were selected to compose two groups of 10 pigs showing extreme and divergent characteristics for the BFT EBV with respect to the larger population of the 949 pigs (Table 1). The twenty animals considered for RNA-seq analysis were slaughtered in 12 dates, with 5 dates common to both groups. The animals were selected also according to their pedigree, in order to avoid the presence of full sibs in the considered groups. From now on the two groups will be referred as FAT and LEAN samples.

RNA extraction, library preparation, sequencing

Total RNA was extracted with Trizol (Invitrogen) according to the manufacturer's instruction. RNA extracted samples were quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies) and the quality of the RNA was assayed using an Agilent 2100 BioAnalyzer (Agilent Technologies). The RNA libraries were prepared from total RNA using the TruSeq RNA sample preparation kit (Illumina) and version 3 of the reagents, following the manufacturer's suggested protocol. The libraries were tagged and couples of libraries were run on a single lane of an Illumina HiSeq2000. Reads are 100 nt paired-end represented in FASTQ format.

Architecture of the bioinformatics pipeline

A computational pipeline to process the sequencing data for gene/transcript expression estimation and to perform differential expression analysis between the two sample groups was developed. The pipeline components to achieve expression estimates were assembled

using Scons software (http://www.scons.org/. Accessed 22 June 2015), which allows the parallelization and automation of the pipeline tasks. The pipeline and its following steps are detailed in the next paragraphs.

RNA-seq data pre-processing and mapping to swine genome

Exploratory analyses on the raw reads quality were carried out using the FastQC v0.10.1 software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 22 June 2015), which generates an HTML report for each sample read set. Read fragments with quality Phred score lower than 30 were trimmed using the DynamicTrim script of the SolexaQA v2.1 (Cox et al., 2010). The FASTX-Toolkit v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/. Accessed 22 June 2015) was used for trimming result report. A custom Python script using the HTSeq package (Anders et al., 2015) filtered out the trimmed reads shorter than 50 nucleotides. To maintain a consistent paired-end read set, discarded read mates were also filtered out, despite their length and quality. Each sample paired-end clean read set was mapped to the swine genome (Sscrofa10.2.70) by Tophat v2.0.8 (Kim et al., 2013) using default parameters with transcriptome inference from Ensembl annotation (Tophat2 used Bowtie v2.1.0.0; Langmead & Salzberg, 2012) and SAMtools v0.1.19()Li et al., 2009).

Gene/transcript expression evaluation and transcript reconstruction

Gene annotation for the reference genome was retrieved from Ensembl (BioMart) using the biomaRt R package (Durinck et al., 2009). Read alignments were processed by Cufflinks v2.1.1 (Roberts et al., 2011a; Roberts et al., 2011b; Trapnell et al., 2010) to identify and

discover expressed genes and transcripts, and to quantify their expression. Expression data were indicated as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Cufflinks was applied to each sample alignment; then, we merged the transcript predictions in a non-redundant reference using the Cuffmerge tool from the Cufflinks package. To reduce artefacts deriving from the transcript prediction and normalisation strategies, only predicted transcripts at least 200 nt long and with minimal expression of 100 (Cufflinks normalised) reads in at least one of the two groups were considered for transcriptome reconstruction and for the following analyses.

Gene and transcript differential expression assessment

The samples were inspected by principal component analysis to examine their similarities. The read counts of each gene in the 20 considered samples were transformed with the variance stabilizing transformation function provided by the DESeq2 package (Anders & Huber, 2010) and used to compute the principal components.

The genes identified by Cufflinks were assessed for differential expression (DE) between the LEAN and FAT groups, by means of two strategies, namely Cuffdiff2 (v2.1.1 from the Cufflinks package; Trapnell et al., 2012) and DESeq2 v1.2.1 (Anders & Huber, 2010). Instead, transcript DE was assayed only with Cuffdiff2. To represent gene expression, the two methods use similar statistical approaches based on generalized linear model (GLM) of the negative binomial family. Cuffdiff2 extends the model using a beta negative binomial distribution to handle uncertainty of multi-mapped reads. On the contrary, DESeq2 considers only uniquely mapped reads (counted by means of the htseq-count script of the HTSeq package (Anders et al., 2015), but facilitate the specification in the statistical model

of additional factors effecting the fit of the GLM. In this study, the statistical model included sex effect as a potential conditioning factor. Gene and transcript DE test computed P-values were corrected according to the Benjamini-Hochberg procedure. Differentially expressed genes and transcripts were considered statistically significant according to false discovery rate less than or equal to 0.05.

Transcript characterisation

Using custom scripts including BEDTools v2.17.0 software (Quinlan & Hall, 2010), we retrieved the nucleotide sequences of the transcripts extracting from the *Sus scrofa* genome the stretches of nucleotides according to the annotation generated by the RNA-seq analysis tools. Transcripts were identified or characterised by sequence similarity using BLASTN and BLAST2 from the NCBI BLASTN suite (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Accessed 22 June 2015) using Megablast algorithm (Morgulis et al., 2008). To assign a gene name, the sequences IDs obtained with this comparison were used to query the NCBI Gene and the UniGene databases (http://www.ncbi.nlm.nih.gov/unigene/. Accessed 22 June 2015). We used two strategies for transcript annotation. DE transcripts and genes were annotated by similarity using nr/nt nucleotide collection. The threshold considered for the identification of our transcripts was identity ≥80% in at least 70% of the sequence length of a transcript present in the database. Transcripts from new genes were characterized using a comparative genomics approach. We compared the new transcripts from intergenic regions with known human transcripts (RefSeq Release 72) by aligning with BLASTN (NCBI BLAST 2.2.29+). For each transcript the best hit was considered, and then

alignments with E-value greater than 10e-6, identity less than 60%, and length less than 100 nucleotides were discarded.

Prediction of coding/non-coding potential

The transcript coding potential was predicted by CPC (Coding Potential Calculator; Kong et al., 2007). CPC is a support vector machine-based classifier of transcript protein-coding potential grounding on six features of sequence. Three features assess the extent and quality of the predicted transcript ORF: the Framefinder software identifies the longest ORF in the three forward and in the three reverse frames, then the coverage and the integrity of the predicted ORF are evaluated. Another three features derive from results of BLASTX search against UniProt Reference Clusters. All the features contribute together to a final score, and to the classification of transcripts as coding or non-coding. Only transcripts not including uncalled bases were considered for CPC analysis.

Validation by quantitative real time-PCR

The validation of selected RNA-seq results was performed using a quantitative real time-PCR (qRT-PCR) approach using 18 out of the 20 samples used for the RNA-seq analysis. Two samples, one in the FAT group and one in the LEAN group, were not considered because the total RNA extracted was used completely for the RNA-seq analysis. QRT-PCR validation was carried out using Rotor-Gene TM 6000 (Qiagen - Corbett Research). After DNase treatment (TURBO DNA-free^{TM}, Ambion, Applied Biosystems), 1 µg of total RNA was reverse transcribed using the iScript cDNA Synthesis kit (BIORAD) according to the manufacturers' instructions.

The samples were first used to analyze four candidate normalizing genes beta-2-microglobulin (*B2M*), polymerase (RNA) II (DNA directed) polypeptide A, 220kDa (*POLR2A*), hypoxanthine phosphoribosyltransferase 1 (*HPRT1*), tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta (*YWHAZ*). The primer pairs and the PCR conditions used are reported in Supplementary Table 1. The expression levels of these four genes were evaluated using NormFinder and *B2M* and *HPRT1*, the two most stably expressed normalizing genes, were utilized as reference genes. For each gene selected for validation, we designed an external primer pair to obtain the amplicon for the standard curve construction and an internal primer pair for the qRT-PCR on Rotor Gene 6000 (Table S1). Standard curves for each gene were generated from 10-12 serial dilutions (from 10^9 to 25 molecules/µl) of the PCR amplicons obtained with the external primer pairs and containing the internal primers used in the qRT-PCR analysis. Amplifications were performed in a total volume of 10 µl containing using 5 µl of the SYBR® Premix Ex Taq™ (Takara Bio Inc.), 0.5 µl of each primer and about 100 ng of cDNA. The used Premix Ex Taq™ is optimized for a two-step cycling, and the amplification conditions for the tested genes are reported in Table S1. The PCR efficiency was calculated as E=10 exp(-1/slope), with a range between -2.7 and -4.3, indicating a good PCR efficiency result. All the PCR products were checked on a polyacrylamide gel and the specificity of the amplification was checked by a final melting curve analysis.

Threshold cycles obtained for the samples were converted by Rotor Gene 6000 to mRNA molecules/µl using for each gene the relative standard curve (Bustin & Nolan, 2004). Moreover, the average mRNA molecules/µl for each sample was normalized dividing the mRNA molecules of a gene /µl by the geometric average of *B2M* and *HPRT1* mRNA molecules/µl in the given sample, as suggested by Bustin & Nolan, 2004 and Vandesompele

et al., 2002. Differences on the expression level calculated for FAT and LEAN samples were tested by two-tailed Student's t test. Statistical analyses were performed with SAS version 9.3 (SAS 9.3 Help and Documentation, Cary, NC. SAS Institute Inc.) and nominal P-value ≤0.05 was considered as significance threshold.

Functional characterization

Functional annotation, classification and annotation clustering of selected gene sets were carried out by DAVID Tools 6.7 (Huang et al., 2009) using Biological Processes, Molecular Function gene ontology categories and KEGG pathways. A threshold for significance of P<0.01 and P<0.05 after Benjamini correction was considered for the selection of the functional categories respectively in the characterization of most expressed transcripts and for the selection of the functional categories of DEG.

Results

Samples

In this study we applied RNA-seq by Illumina technology to the study of gene expression in backfat tissue of 20 ILW pigs. We considered a large group of 949 sampled animals, with EBV for BFT ranging from -10.64 mm to 7.28 mm, with mean value and standard deviation (SD) -1.96 mm and 3.01, respectively. We selected, from the whole collected population, two groups of 10 unrelated pigs (FAT and LEAN) with extremely divergent EBVs for BFT, with 1:1 sex ratio within each group. The mean values of each of the two selected groups of pigs are outside the range -7.98 mm / 4.06 mm defined by the mean value of the 949 samples ±2

standard deviations. Specifically, FAT and LEAN animals were associated to average BFT values of +5.22 mm (±1.30 SD) and -8.63 mm (±1.40 SD) as indicated in Table 1.

Sequencing, reads pre-processing and mapping

Pairs of samples were run together, after barcoding, on a single lane of an Illumina HiSeq 2000 apparatus, obtaining a total of 3,917,123,414 raw reads for the 20 considered samples, with an average of 195,856,171 raw reads per sample (Table S2; GEO accession GSE68007). After trimming and length filtering the clean reads per sample were on the average 113,934,264 (58.04%) and were used for read-to-genome mapping (Figure S1A). Reads that align on a single genome locus (uniquely mapped reads) were on the average the 91.07% of the mapped reads (Table S2). The 72.42% of the uniquely mapped reads (72,219,306.45 on the average aligned to annotated exons, the 19.15% mapped on intergenic regions and the 8.43% mapped on introns of annotated genes. The deep sequencing allowed the identification of genes expressed at low level and relatively rare alternatively spliced transcripts. We observed splicing events in the 21.19% of the reads on the average, providing useful information for the reconstruction of alternative transcript isoforms (Figure S1B).

Transcripts and genes expressed in backfat samples

The deep sequencing analysis of backfat transcripts performed on two groups of pigs divergent for fat deposition in this tissue allowed the detection of 63,418 transcripts. Many of them have not yet been annotated in the porcine genome, thus providing new consistent resources for pig genome annotation and studies of adipose tissue biology. We identified

the expression of genes on all porcine autosomes, sex chromosomes and mitochondrial genome. Chromosome 1 has the largest number of expressed genes (8.23%), followed by chromosomes 6 (7.84%) and 2 (7.25%). Furthermore, a non-negligible part (12.48%) of the expressed genes is located in genomic scaffolds (Figure S1C), as about the 7.5% of the genome has no assigned location yet, as described in Ensembl annotation of pig genome (database version 78 at the time of the analysis ; http://www.ensembl.org/Sus_scrofa/Location/Genome. Accessed 22 June 2015). In term of genes, we identified 23,483 expressed pig genes: 12,707 known and 10,776 putative new genes.

Transcripts were split in different classes according to their matching with the genome annotations (Figure 1A, Table S3). Transcripts matching exactly the reference annotation are indicated as “known” transcripts; annotated transcripts’ new isoforms or overlapping with annotated transcript are indicated as “novel isoforms; and all other transcripts, such as those expressed from extragenic regions, are referred as “new” transcripts and might represent putative new genes. The majority of expressed transcripts are novel isoforms (35,030; the 55.2%) or known transcripts (12,969, representing the 20.5%) that are prevalently annotated as protein coding (12,883; 99.3%). The expressed new transcripts are 15,419 (24.3%).

Transcript lengths range from 200 to 50,610 nt, with median and average values of 3,224 and 3,979. Average size exceeds the 2 kb pig mean transcript size that can be estimated according to Ensembl pig coding transcript annotation. We observed that the novel isoforms reconstructed are longer than “known” pig transcripts (Figure 1B).

Sequences longer than 5 kb compose the 25% of the expressed transcripts. Noteworthy, we detected two transcripts overlapping *ZBTB16* gene and two new transcripts from chromosome 16 that are longer than 40 kb.

Considering transcripts expression, we observed that new transcripts are less expressed in fat tissue than known transcripts (Figure 1C). Nevertheless, all the three transcript categories span a considerably large range of expression values.

The majority of the expressed genes (12,138; 52%) present only one transcript isoform expressed in fat tissue (Figure 1D); the 27.0% and the 18.3% of the genes present two and three expressed isoforms, respectively, whereas the remaining 12.7% of the genes are associated each one to 4 to 31 different isoforms. We identified 31 isoforms for the gene *MAP4K4*, for which a complex expression pattern is reported in humans: Ensembl release 79 lists 20 *MAP4K4* transcripts generated by at least 3 different promoters, by complex alternative splicing and by polyadenylation patterns, whereas five protein isoforms are reported in UniProt release 2015_3.

Looking at isoform types, Figure 1E shows that many genes expressing only one transcript (first bar from the left) in fat tissue are putative new genes (green portion). Interestingly, some genes expressing only one transcript in fat tissue are represented only by a novel isoform (first bar, red shadows). The proportion of novel isoforms (red portion) increases along with the numbers of expressed transcripts per gene. Moreover, the transcripts classes showing exonic overlap compared to a reference transcript are found in genes with a varying number of transcripts and are particularly abundant in genes with up to three isoforms. The remaining transcript classes are very rare.

Interesting new isoforms derived from known genes regard Perilipin 2 (*PLIN2;* alias *ADFP*, adipofilin), an important gene for fat metabolism in pigs (Davoli et al., 2010; Gandolfi et al.,

2011) whose expression in humans correlates positively with cytosolic triacylglicerol levels (Conte et al., 2013). Only one transcript is currently annotated in Ensembl for pig *PLIN2* (ENSSSCT00000005701), whereas according to our results, *PLIN2* expressed four different isoforms. The most expressed *PLIN2* transcript (expressed two times more in FAT than in LEAN pigs) is a non-annotated isoform (TCONS_00002441 in Table 2; 2441DE in Figure S3) characterized by the skipping of the fourth exon. The same transcript has also a shorter 3' sequence with respect to the canonical *PLIN2/ADFP* form, probably due to the use of an alternative polyadenylation site. Importantly, the skipping of the 83 nt long exon four introduces downstream a shift in the reading frame and a premature stop codon. Thus, this transcript encodes a truncated protein (only 80 aa) corresponding to the N-terminal region and of the Perilipin domain of the *PLIN2* protein annotated isoform (463 aa). The other two new transcripts differ from the annotated isoform, one for the skipping of exon 2, and the other for a longer first exon, probably due to alternative TSS usage by different promoters. The four expressed isoforms are also heterogeneous in the length of the 3' UTR region.

Coding and non-coding transcripts from new genes

We obtained a characterization of intergenic transcripts from new genes first both by similarity, comparing them against human transcripts, and by predicting their coding potential. New pig transcripts with an assigned human best hit were 10,020 (65%), expressed by 7,099 genes (66%), and corresponding to 4633 human Refseq sequences (3,882 unique gene symbols; Table S4).

We considered 12,702 intergenic transcripts for protein coding potential analysis. For each transcript, the coding potential of both the forward and the reverse complement sequence

were evaluated. According to CPC results, we classified the 35.8% (4,551) of transcripts as coding, and the 64.2% (8,151) as non-coding. As done by Zhou et al., (2014), we considered as proper non-coding only those transcripts classified as non-coding and having a CPC score lower than -1 for both the forward and the reverse sequence. A portion of the non-coding transcripts (37.5%) resulted with CPC score < -1 for both the forward and the reverse complement sequences. We refer to these transcripts as "reliable non-coding" class, which represented 24% (3,056) of the intergenic transcripts (Figure 2A). We observed that intergenic coding transcripts are on average longer than intergenic non-coding transcripts (4,149 and 3,083 nt, respectively), and that the reliable non-coding fraction has a even shorter average length (2,571 nt; Figure 2B and Table S5). Reportedly, non-coding transcripts tend to be shorter and to have fewer exons than coding transcripts in mammalian genomes (Iyer et al., 2015).

Coding transcripts have an average expression in fat tissue higher than the non-coding transcripts (5.32 and 2.28 FPKM respectively, and 3.23 FPKM for the reliable non-coding group; Figure 2C). One reliable non-coding transcript is ranked within the 100 most expressed transcripts detected in backfat tissue; 15 reliable non-coding transcripts are within the 1,000 most expressed transcripts; and 98 are within the 10% most expressed transcripts (Table S6). In agreement with previous results showing that coding transcripts tend to present higher expression than non-coding ones (Cabili et al., 2011; Iyer et al., 2015), we observe that intergenic transcripts ranking in the 10% most expressed in backfat tissue are enriched in the coding category (55%) and particularly if compared with the proportion of the coding category within the set of intergenic transcripts (35.8%; Figure 2D, green portions).

Function of most expressed transcripts

A global view of the transcription profile of porcine backfat tissue was obtained by averaging the FPKM values of all 20 analysed samples. The 1411 most expressed transcripts, accounting together for 75% of expression, were chosen to extract the most expressed genes (Table S6).

Among these genes, 59 are indicated as reliable non-coding (CPC score <1) and 66 showing a positive CPC score are indicated as putative coding.

According to DAVID functional annotation and clustering, we characterized the biological processes (Table S7) associated to the most expressed genes. Ribosomal activity, oxidative phosphorylation, protein metabolic processes, intracellular protein transport, regulation of translation initiation, fatty acid metabolism, response to oxidative stress resulted to be the biological processes more represented in subcutaneous adipose tissue of the analysed samples.

Gene/transcript differential expression

Unsupervised analysis of gene expression profiles was carried out to inspect similarities among the samples. Principal component analysis revealed a clear separation of the LEAN and FAT samples according to the first two most informative components (Figure S4 A), which, notably, do not separate the samples by sex (Figure S4 B).

Average gene expression values for FAT and LEAN groups were 32.46 and 33.63 FPKM). In both groups, few highly expressed genes contribute to the majority of the cumulative expression. For instance, roughly 25% of expressed genes (5,908 and 5,728 in FAT and LEAN, respectively) constitute 95% of the total detected expression (Figure S2). As expected,

transcript expression distribution is similar to the gene expression distribution being positively skewed, with mean and median corresponding to 11.84 and 0.64 FPKM, respectively. Transcripts average expression values are lower than genes expression values since the latter was computed as the sum of transcripts expression of each gene.

To identify a set of robust DEG and DET the transcription profiles of FAT and LEAN samples were compared with the integration of two methods applied at gene and at transcript levels. Cuffdiff2 identified 414 DEGs between FAT and LEAN groups, corresponding to 1,187 transcripts: 266 DEGs are more expressed and 148 DEGs are less expressed in FAT samples. Fold changes in base two logarithmic scale of DEGs range from 0.46 to 8.95 for the higher expressed genes, and from -6.19 to -0.47 for the less expressed ones (Table S8). DESeq2 identified 586 DEGs (185 in common with the DEGs identified by Cuffdiff2) corresponding to 1,504 transcripts: 358 genes are up-regulated and 228 genes are less expressed in FAT samples. DEGs base two logarithmic scale transformed fold changes (Log_2 FC) range from -1.13 to -0.20 for the less expressed genes and from 0.21 to 1.18 for the higher expressed genes (Table S9). Cuffdiff2 differential expression analysis at the transcript-level identified 154 DE transcripts (corresponding to 153 genes): 48 were less expressed and 106 transcripts were more expressed in FAT samples, with Log_2 FC ranging from -3.44 to -0.54 and from 0.64 to 3.66, respectively (Table S10). On the whole, 818 genes were DE, or associated to at least one DE transcript, according to at least one method, were detected (Figure 3A).

The overlapping of the different lists of DEGs and the list of DE transcripts (DET) evidenced a group of 86 DET that are identified by all the approaches, from now on referred as "common DET" (cDET). These DET belongs to 78 DEG, from now on referred as "common DEG" (cDEG) since five genes are represented by more than one isoform (Table 2).

The cDET present the same fold change sign of the corresponding cDEG (Figure 3B): 72 DET were more expressed in FAT (max Cuffdiff2 gene-level Log_2 FC 2.55 for *DSC2* gene) and 14 DET were less expressed in FAT (minimum Log_2 FC -3.44 for an intergenic gene located in GL894890.1 scaffold). Among the 86 cDET, 44 are known transcripts, 16 are novel isoforms and 26 come from intergenic regions.

cDEG are found in all chromosomes except for chromosomes 16 and Y, with up to 11 DE genes in chromosome 4 and 19 DE genes in scaffolds (Figure 3C). The most expressed (average FPKM greater than 100) known cDEG, reported in decreasing FPKM order, are *DNAJB1*, *CTSH*, *CTGF*, *C1QC*, *SPP1* and *CDO1*.

Coding and con-coding intergenic DET

We considered the 41 novel isoforms or new transcript cDET for CPC analysis. In 14 of these transcripts both the forward and reverse sequence is probably non-coding, according to integrated ORF analyses and to similarity searches, and to CPC score thresholds used before. Five cDET with CPC score <-1 were scored as “reliable non-coding”. Of the remaining transcripts, nine presented low coding potential both in the forward and in the reverse complement sequence but with CPC score ranging from -1 to 0 (“non-coding”), and 27 were classified as coding transcripts (Table S11).

qRT-PCR confirmation of DE for selected genes

To validate the results obtained by RNA-seq, eleven cDEG were chosen according to the absolute value of the Log_2 FC between FAT and LEAN pigs, or for their functional role and involvement in relevant pathways. As reported in Figure 4, the DE of all selected genes has

been validated, with high correlation between the fold changes obtained by RNA-seq and by qRT-PCR data.

DE transcript characterisation

We characterized the cDEG in terms of their functional role in adipose tissue. Using DAVID Bioinformatics Resources we first identified the functional categories, enriched in genes differentially regulated between FAT and LEAN groups.

The Biological Process categories enriched in higher expressed DEG are response to stimulus, immune system process and cell activation, skeletal system development (Table 3). DAVID clustering of the few lower expressed genes detected (*ADSSL1, CDO1, DNAJB1, HSPA1A, HSPA1B, HSPA2, HSPB8, IGFBP5, OLFML3*) allowed to identify the functional categories unfolded protein binding and stress response represented by five heat shock protein genes that are involved in protein stabilization after cellular stress.

Apart from the Gene Ontology-based functional characterization of the whole subsets of higher- and lower-expressed genes we considered cDEG function and involvement in specific pathways, according to literature and knowledge bases.

Several more expressed genes in FAT animals (*ACP5, BCL2A1, CD1A, EGR2*, *ENPP1*, *GPNMB*, *INHBB, LYZ, MSR1, OLR1, PIK3AP1, PLIN2, SPP1*, *STC1*) are characterized by a metabolic function mainly related to adipocyte growth regulation, while others (*CCR1, CD163*, *SLC11A1*) are known to be involved in immune defence of the organism.

Discussion

Transcriptome data highlight the adipose tissue complexity

The deep sequencing analysis of pig backfat transcriptome performed allowed finding thousands of genes and transcripts expressed. In the present study, we applied stringent cleaning and filtering procedures of the sequencing data and, on average, 90 million reads per sample were mapped, obtaining a higher sequencing depth compared to previous studies (Chen et al., 2011; Jiang et al., 2013; Sodhi et al., 2014; Wang et al., 2014).

The adipose tissue is not only metabolically and transcriptionally active, but has been recognized as an important endocrine organ (Kershaw et al., 2004; Trayhurn et al., 2005). Adipocytes are a dynamic and highly regulated population of cells (Rosen & MacDougald, 2006; Moreno-Navarrete & Fernández-Real, 2012). Our results agree with these data supporting the characterization of the adipocytes as highly specialised endocrine cells that can play key roles in various physiological processes. The multifunctionality and the complexity of the tissue is witnessed also by the high number of transcripts (more than sixty thousands) found in the present study, including many new transcripts from previously non-annotated loci in porcine genome. The majority of the reconstructed sequences are novel isoforms of already known genes that express more than two different transcripts each. Similar patterns observed in human cells (Djebali et al., 2012) and the high quality of the sequenced reads used in our analysis support the idea that this is more attributable to an incomplete annotation of the transcript isoforms expressed in pig backfat, than to transcript reconstruction artefacts. The different isoforms derived from the same locus arisen from our analysis and observed for almost half of the expressed genes, may contribute to improve the knowledge of the porcine transcriptome, and to refine the current swine genome annotation. The new *PLIN2* isoforms reported above are an interesting example, especially if compared to the human genome where at least eight *PLIN2* transcript isoforms are annotated and only four of them are coding. Remarkably, three human *PLIN2* isoforms

encode N-terminal truncated amino acid chains that are similar to the truncated isoform we reconstructed in our study, and whose function has not yet been elucidated. Furthermore, Russell et al. (2008) identified in a *PLIN2* deficient mouse cell line the expression of a *PLIN2* C-terminal truncated protein that may partially replace the function of the full-length protein. Additional studies are needed to understand if and how the short transcript we found differentially expressed could change the gene functions compared to the wild type long protein.

Functional characterisation of the adipose tissue expression profile

The profile of the subcutaneous adipose tissue transcriptome in pigs was delineated and the functional analysis of the genes expressed in backfat tissue was performed to know their metabolic role and to connect them to specific competences of the tissue. We didn't find particular differences between the functional categories of the genes expressed in the backfat tissue of FAT and LEAN pigs. More in details among the most expressed genes in the fat tissue, many are involved in metabolic pathways and biological processes related to protein metabolism, oxidoreductase activity for ATP production, regulation of lipid synthesis and degradation.

Genes differentially expressed between LEAN and FAT animals converge and connect to specific functions

The detection of DE genes and transcripts has been obtained by a stringent procedure grounding on integration of different methods for expression estimation and differential expression testing, as done in a recent study (Ropka-Molik et al., 2014) focused to muscle

tissue gene expression in pigs of different breeds. In the present study, which compares pigs of the same breed and reared under standard conditions, we detected significant gene expression variations. The sensitivity of our approach was supported by the successful validation of all the eleven DEG assayed.

We analyzed the biological functions of genes differentially expressed between FAT and LEAN animals (Figure 5). It is interesting to note that the main differences were found for functional categories of genes related Inflammation and immunity that resulted more expressed in FAT pigs. The genes less expressed in FAT animals include some heat shock protein genes. The biological functions of DEGs show a stronger activation in adipose tissue of FAT pigs of genes for important processes involved in hypertrophy and adipogenesis, such as differentiation and maturation. Supposedly, these biological processes could be altered in adipose tissue of FAT pigs due to dysregulated adipose metabolism and endocrinology similarly to what was hypothesized in humans (Sethi, 2010). On the whole, there is a consistent difference concerning the biological functions characterizing the most expressed genes on backfat tissue and those of the genes differentially expressed between FAT and LEAN pigs.

Some genes higher expressed in FAT animals could modulate backfat physiological processes

Specific DEGs more expressed in FAT pigs participate to biochemical pathways related to and involved in adipocytes metabolism and adipose tissue physiology. Ectonucleotide pyrophosphatase/phosphodiesterase 1 (*ENPP1*) encodes a catalytic enzyme involved in adipocyte maturation (Liang et al., 2007). Pan et al. (2011) showed that the over-expression

of *ENPP1* in a human cell line resulted in adipocyte insulin resistance and demonstrated an association with fatty liver, hyperlipidemia, and dysglycemia. Accordingly, the study of Chandalia et al. (2012) underlined an increased *ENPP1* expression in adipose tissue associated with defective adipocyte maturation leading to pathogenesis of insulin resistance and its associated complications for glucose and lipid metabolism in absence of obesity. In addition, Meyre et al. (2005) reported the presence of three *ENPP1* SNPs in human gene associated with adult obesity and increased risk of glucose intolerance and type 2 diabetes. Furthermore, also the genes acid phosphatase 5, tartrate resistant (*ACP5*) and lysozyme (*LYZ*) that in this research have higher transcriptional level in FAT pigs have been reported to be involved in excessive backfat deposition in pigs and in the development of atherosclerosis (Padilla et al., 2013).

In the present research, some genes overexpressed in the adipose tissue of FAT pigs, namely *STC1, EGR2,* and *INHBB*, are related to adipocyte differentiation and adipocyte maturation. *STC1* (Stanniocalcin 1) has been reported in literature to be up-regulated during adipogenesis and to modulate steroidogenesis. Serlachius & Andersson (2004) related *STC1* up-regulation to the set of survival genes in adipocyte differentiation, which is also associated to overexpression of the anti-apoptotic proteins *BCL2* reported to be involved in inflammation pathway. *EGR2* (early growth response 2) is a direct target of *mir-224-5p*, a negative regulator of adipocyte differentiation that is down regulated during the early process of mouse adipocyte differentiation, and the expression of *EGR2* is increased (Peng et al., 2013). The *INHBB* (Inhibin beta B) gene coding for the activin B subunit is part of the inhibins/activins family of proteins with cytokine and hormone activity. In human and mice, INHBB has been associated to the physiological and metabolic modifications during adipogenesis when it is highly expressed and is the predominant activin in human adipose

tissue (Hoggard et al., 2009). *INHBB* is member of TGF-protein superfamily of secreted growth factors involved in many biological responses including regulation of apoptosis; proliferation and differentiation of human adipocytes; tissue remodeling; and inflammatory immune response (Dani C., 2013). It can be hypothesized that in FAT pigs the pro-adipogenic *INHBB* gene expression increases as it is involved in the differentiation of preadipocytes into mature adipocyte, and that *INHBB* is involved in many physiological processes and including the control of food intake and to energy metabolism through the regulation of hypothalamic and pituitary hormone secretions. Another gene overexpressed in FAT pigs related to feeding and pituitary secretions is *GPNMB* (glycoprotein transmembrane NMB). GPNMB is one of the receptors activated by bombesin-like endogenous peptide ligands, such as gastrin-releasing peptide (*GRP*), neuromedin B (*NMB*) and neuromedin C (*GRP18-27*). These receptors are involved in the regulation of many biological functions including thermoregulation, feeding, pituitary, gastric and pancreatic secretion. The NMB/NMB-R pathway is involved in the regulation of a wide variety of behaviours, such as spontaneous activity, feeding, and anxiety-related behaviour (Yamada et al., 2002).

The *OLR1* (Oxidized low density lipoprotein (lectin-like) receptor 1) gene resulted more expressed in FAT pigs compared to LEAN animals. This gene codes for a LDL receptor that belongs to the C-type lectin superfamily, one of many target genes, including perilipins, of the PPAR signalling, which is involved specifically in lipid metabolism and fatty acids transport. In this way, *OLR1* is a receptor that mediates the recognition, internalization and degradation of oxidatively modified low-density lipoprotein by vascular endothelial cells. *OLR1* removes oxidised low-density lipoproteins from the circulation, as part of lipid metabolism pathways (Mehta et al., 2002).

Genes involved in immunity and inflammation are more expressed in FAT animals

Some other genes overexpressed in FAT pigs are related to immunity. Inflammatory links between human obesity and metabolic diseases are well known mechanisms based on the recruitment of immune cells into adipose tissue (Kabir et al., 2014). The development of a pre-inflammatory condition in presence of dysregulated excessive adipogenesis is associated with adipose macrophage infiltration and activation. From our study, we can hypothesize a similar process in backfat tissue of FAT pigs where we identified the over expression of the gene macrophage scavenger receptor 1 (*MSR1*), a membrane glycoprotein that in humans is involved in the pathologic deposition of cholesterol in arterial walls during atherogenesis (Haasken et al., 2013). Additionally, the overexpression of secreted phosphoprotein 1 (*SPP1*) in FAT pigs can suggest the hypothesis that this gene is acting as a proinflammatory cytokine that promotes monocyte chemotaxis and cell motility and might link, in pigs like in mice, fat accumulation to the development of insulin resistance by sustaining inflammation and the accumulation of macrophages in adipose tissue (Nomiyana et al. 2007). Interestingly, a porcine *SPP1* gene polymorphism was associated to backfat thickness in the Landrace × Jeju (Korea) Black pig F2 population (Han et al., 2012). *SPP1* might play a key role in the pathway that leads to type I immunity enhancing interferon-gamma and interleukin-12 production and suppressing interleukin-10 (Ashkar et al., 2000). Therefore, these data allow hypothesizing *SPP1* as a gene associated, in pigs like to in human, to the link between obesity, adipose tissue inflammation, and insulin resistance. In addition, phosphoinositide-3-kinase adaptor protein 1 (*PIK3AP1*), higher expressed in FAT pigs, is a positive regulator of phosphatidylinositol 3-kinase (*PI3K*) signalling. *PI3K* signalling pathway

has a key role in the insulin-dependent regulation of adipocyte metabolism (glucose and lipid metabolism). Besides, *PI3K* participate in obesity-associated inflammatory cell recruitment (neutrophils and macrophages), as well as in the CNS-dependent neurohumoral regulation of food intake/energy expenditure (McCurdy & Clemm, 2013; Beretta et al., 2015).

Other genes found in the present research and related to inflammatory condition of the adipose tissue in FAT pigs are particularly interesting to mention. CD163, member of the scavenger receptor cysteine-rich superfamily (Guo et al., 2014; Smith et al., 2014); solute carrier family 11 (proton-coupled divalent metal ion transporter), member 1 (*SLC11A1*), a gene involved in the resistance to *Salmonella* infection (Kommadath et al., 2014) as well as the chemokine (C-C motif) receptor 1 (*CCR1*), that was previously found overexpressed in obese pigs (Kogelman et al., 2014); BCL2-related protein A1 (*BCL2A1*), a gene found to be overexpressed in pigs with an high obesity index and that is related to immunity, inflammatory pathway, and osteoclast differentiation (Kogelman et al., 2014); CD1a molecule (*CD1A,* indicated as *PCD1A* on the cited paper), a surface antigen involved in immunity was found to be overexpressed in obese pigs by Kogelman et al. (2014). The same Authors highlighted a strong connection between fat deposition on the body (obesity), immunity and bone development. They also indicated that *CCR1* gene is a strong candidate regulator of immune response as it is a receptor of pro-inflammatory chemokines in adipose tissue playing a pivotal role in obesity-associated diseases (Kabir et al. 2014; Lumeng & Saltiel, 2011).

Heat shock response , protein folding and repair are impaired in FAT animals

Considering the 14 genes less expressed in FAT animals, direct relationships with lipid metabolism are not apparent. However, the “unfolded protein binding” function is enriched among these genes, which include five functionally linked heat shock proteins (*DNAJB1, HSPA1A, HSPA1B, HSPA2* and *HSPB8*). Heat shock proteins are involved in stabilization of existing proteins against aggregation, mediating the folding of newly translated proteins in the cytosol and in organelles, and also in the ubiquitin-proteasome pathway. *DNAJB1*, a member of the Hsp40 family, is a molecular chaperon involved in protein folding and protein complex assembly. *DNAJB1,* a member of the Hsp40 family, promotes protein folding and prevent misfolded protein aggregation, as *HSPB8*, a member of the Hsp20 family, does (Vicario et al., 2014). *DNAJB1* also stimulates the ATPase activity of protein of the Hsp70 family to which other genes less expressed in FAT pigs (*HSPA1A*, *HSPA1B*, and *HSPA2)* belong, indicating a possible functional link between these four genes. Our results suggest a general impairment of the protein folding and repair in the fattest animals, in accordance to previous observations of studies carried out on human obesity. Obesity is a pathological human condition in which a chronically positive energy balance induces in adipocytes, the cells in charge to store the excess of energy in fat depots, a persistent stress activating in turn defence processes as autophagy or apoptosis.

As reviewed by Newsholme & de Bittencourt (2014), if the heat shock response, a key component of the physiological response to resolve inflammation, is hampered in adipose tissue, the adipocyte metabolic stress triggers fat cell senescence with reduction of the heat shock proteins activity. In this condition, the advance of inflammasome mediated secretory activity from adipose to other tissues promotes cellular senescence in many other cells of the organism, aggravating obesity-dependent chronic inflammation. This mechanism could have been activated also in the FAT pigs of our experiment (Figure 5) due to a genetic

aptitude of the fattest animals toward a higher fat deposition and adiposity similar to obesity. Indeed, a decrease in the synthesis of the mRNAs of the heat shock proteins and an increase of the expression of many genes related to an inflammatory status and to immune response is a characteristic of the fattest pigs. Increase of the expression of *INHBB* and *SPP1* denotes for instance the augmented production of cytokines and the higher expression of *ENPP1* and *PIK3AP1* may indicate a status of insulin resistance, one of the typical signals connected with obesity.

Pig backfat deposition and impaired stress response may activate inflammation

Our results agree with recent studies showing that several immune system and anti-inflammatory processes are activated and play a critical role in the response to fat accumulation in porcine backfat tissue (Sodhi et al., 2014) and in visceral fat tissue (Toedebusch et al., 2014; Wang et al., 2014). Wang et al. (2014) and Zhou et al. (2013) used three female Landrace pigs to identify DEG between subcutaneous, visceral and intramuscular fat indicating that visceral and intramuscular adipose tissues were mainly associated with inflammatory features of the tissue and immune response. Our data suggest that also in backfat a predominant role of immunity processes is related to an increased adipose tissue deposition.

The results obtained seem to sustain the hypothesis that the high fat accumulation in adipose tissue of pigs can determine the development of an inflammatory process producing a cascade of defence and adaptive reactions in the tissue, such as activation of immune system and mesenchymal cells differentiation in adipocytes.

A deeper knowledge of the metabolic processes involved in fat deposition can be very important to develop the use the pig as model species to study obesity and related disorders for humans because of similar anatomy and physiology (Spurlock & Gabler, 2008; Litten-Brown et al., 2010; Varga et al., 2010) and considering the above described similarities between pigs and humans.

In order to fully elucidate the complex gene network regulating backfat deposition on pigs, it will be important to extend the basic knowledge by further coding and non-coding transcriptome characterization. Additional information would probably come from studying interactions between the differentially expressed long RNAs identified in the present paper and the regulatory microRNAs expressed in porcine adipose tissue identified on some of the same animals (Gaffo et al., 2014).

The results of the present work unlock the opportunity that some of the identified differentially expressed genes might be used as biomarkers (Ibáñez-Escriche et al., 2014) to improve carcass fat traits in to look for SNPs regulating their expression to be included in selection schemes to make more sustainable the pig production chain.

Acknowledgements

We thank Dr. Maurizio Gallo from Associazione Nazionale Allevatori Suini (ANAS) and Dr. Luca Buttazzoni (CREA-PCM) for providing the porcine backfat tissue samples, BMR genomics for performing the RNA-Seq. This work was supported by Progetto “AGER - Agroalimentare e ricerca”: Advanced research in genomics and processing technologies for the Italian heavy pig production – Hepiget (Grant N. 2011- 0279).

References

Anders, S., & Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Anders, S., Pyl, P.T. & Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Ashkar, S., Weber, G.F., Panoutsakopoulou, V., Sanchirico, M.E., Jansson, M., Zawaideh, S., Rittling, S.R., Denhardt, D.T., Glimcher, M.J. & Cantor, H. (2000) Eta-1 (osteopontin): an early component of type-1 (cell-mediated) immunity. Science 287, 860–864.

Beretta, M., Bauer, M. & Hirsch, E. (2015) PI3K signaling in the pathogenesis of obesity: The cause and the cure. Adv Biol Regul. 58, 1-15.

Bosi, P. & Russo, V. (2004) The production of the heavy pig for high quality processed products. Italian Journal of Animal Science 3, 309–321.

Bustin, S.A. & Nolan, T. (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 15, 155–166.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. & Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927.

Čandek-Potokar, M. & Škrlep, M. (2012) Factors in pig production that impact the quality of dry-cured ham: a review. Animal 6, 327–338.

Chandalia, M., Davila, H., Pan, W., Szuszkiewicz, M., Tuvdendorj, D., Livingston, E.H. & Abate, N. (2012) Adipose tissue dysfunction in humans: a potential role for the transmembrane protein ENPP1. J. Clin. Endocrinol. Metab. 97, 4663–4672.

Chen, C., Ai, H., Ren, J., Li, W., Li, P., Qiao, R., Ouyang, J., Yang, M., Ma, J. & Huang, L. (2011) A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. BMC Genomics 12, 448.

Chen, C.H., Lin, E.C., Cheng, W.T.K., Sun, H.S., Mersmann, H.J. & Ding, S.T. (2006) Abundantly expressed genes in pig adipose tissue: an expressed sequence tag approach. J. Anim. Sci 84, 2673–2683.

Conte, M., Vasuri, F., Trisolino, G., Bellavista, E., Santoro, A., Degiovanni, A., Martucci, E., D'Errico-Grigioni, A., Caporossi, D., Capri, M., Maier, A.B., Seynnes, O., Barberi, L., Musarò, A., Narici, M.V., Franceschi, C. & Salvioli, S. (2013) Increased Plin2 expression in human skeletal muscle is associated with sarcopenia and muscle weakness. PLoS ONE 8, e73709.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., Pena, R.N., Ballester, M. & Folch, J.M. (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.

Cox, M.P., Peterson, D.A. & Biggs, P.J. (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11, 485.

Dani, C. (2013) Activins in adipogenesis and obesity. Int J Obes (Lond) 37, 163–166.

Davoli, R., Gandolfi, G., Braglia, S., Comella, M., Zambonelli, P., Buttazzoni, L. & Russo, V. (2010) 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.

Durinck, S., Spellman, P.T., Birney, E. & Huber, W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4, 1184–1191.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A.M., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., Gingeras, T.R., (2012) Landscape of transcription in human cells. Nature 489, 101–108.

Ferraz, A.L.J., Ojeda, A., López-Béjar, M., Fernandes, L.T., Castelló, A., Folch, J.M. & Pérez-Enciso, M. (2008) Transcriptome architecture across tissues in the pig. BMC Genomics 9, 173.

Gaffo, E., Zambonelli, P., Bisognin, A., Bortoluzzi, S. & Davoli, R. (2014) miRNome of Italian Large White pig subcutaneous fat tissue: new miRNAs, isomiRs and moRNAs. Anim. Genet. 45, 685–698.

Gandolfi, G., Mazzoni, M., Zambonelli, P., Lalatta-Costerbosa, G., Tronca, A., Russo, V. & Davoli, R. (2011) 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.

Gorodkin, J., Cirera, S., Hedegaard, J., Gilchrist, M.J., Panitz, F., Jørgensen, C., Scheibye-Knudsen, K., Arvin, T., Lumholdt, S., Sawera, M., Green, T., Nielsen, B.J., Havgaard, J.H., Rosenkilde, C., Wang, J., Li, H., Li, R., Liu, B., Hu, S., Dong, W., Li, W., Yu, J., Wang, J., Stærfeldt, H.-H., Wernersson, R., Madsen, L.B., Thomsen, B., Hornshøj, H., Bujie, Z., Wang, X., Wang, X., Bolund, L., Brunak, S., Yang, H., Bendixen, C. & Fredholm, M. (2007) Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags. Genome Biology 8, R45.

Guo, L., Niu, J., Yu, H., Gu, W., Li, R., Luo, X., Huang, M., Tian, Z., Feng, L. & Wang, Y. (2014) Modulation of CD163 expression by metalloprotease ADAM17 regulates porcine reproductive and respiratory syndrome virus entry. J. Virol. 88, 10448–10458.

Haasken, S., Auger, J.L., Taylor, J.J., Hobday, P.M., Goudy, B.D., Titcombe, P.J., Mueller, D.L. & Binstadt, B.A. (2013) Macrophage scavenger receptor 1 (Msr1, SR-A) influences B cell autoimmunity by regulating soluble autoantigen concentration. J. Immunol. 191, 1055–1062.

Han, S.-H., Shin, K.-Y., Lee, S.-S., Ko, M.-S., Oh, H.-S. & Cho, I.-C. (2012) Porcine SPP1 gene polymorphism association with phenotypic traits in the Landrace × Jeju (Korea) Black pig F2 population. Mol. Biol. Rep. 39, 7705–7709.

Henderson, C.R. & Quaas, R.L. (1976) Multiple Trait Evaluation Using Relatives' Records. Journal of Animal Science 43, 1188–1197.

Hoggard, N., Cruickshank, M., Moar, K.M., Barrett, P., Bashir, S. & Miller, J.D.B. (2009) Inhibin betaB expression in murine adipose tissue and its regulation by leptin, insulin and dexamethasone. J. Mol. Endocrinol. 43, 171–177.

Hornshøj, H., Conley, L.N., Hedegaard, J., Sørensen, P., Panitz, F. & Bendixen, C. (2007) Microarray Expression Profiles of 20.000 Genes across 23 Healthy Porcine Tissues. PLoS ONE 2, e1203.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57.

Ibáñez-Escriche, N., Forni, S., Noguera, J.L. & Varona, L. (2014) Genomic information in pig breeding: Science meets industry needs. Livestock Science 166, 94–100.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., Poliakov, A., Cao, X., Dhanasekaran, S.M., Wu, Y.-M., Robinson, D.R., Beer, D.G., Feng, F.Y., Iyer, H.K. & Chinnaiyan, A.M. (2015) The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. 47, 199–208.

Jiang, S., Wei, H., Song, T., Yang, Y., Peng, J. & Jiang, S. (2013) Transcriptome Comparison between Porcine Subcutaneous and Intramuscular Stromal Vascular Cells during Adipogenic Differentiation. PLoS ONE 8, e77094.

Kabir, S.M., Lee, E.-S. & Son, D.-S. (2014) Chemokine network during adipogenesis in 3T3-L1 cells: Differential response between growth and proinflammatory factor in preadipocytes vs. adipocytes. Adipocyte 3, 97–106.

Kershaw, E.E. & Flier, J.S. (2004) Adipose tissue as an endocrine organ. J. Clin. Endocrinol. Metab. 89, 2548–2556.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Kogelman, L.J.A., Cirera, S., Zhernakova, D.V., Fredholm, M., Franke, L. & Kadarmideen, H.N. (2014) Identification of co-expression gene networks, regulatory genes and pathways for

obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Med Genomics 7, 57.

Kommadath, A., Bao, H., Arantes, A.S., Plastow, G.S., Tuggle, C.K., Bearson, S.M.D., Guan, L.L. & Stothard, P. (2014) Gene co-expression network analysis identifies porcine genes associated with variation in Salmonella shedding. BMC Genomics 15, 452.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L. & Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345–349.

Langmead, B. & Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, X.J., Yang, H., Li, G.X., Zhang, G.H., Cheng, J., Guan, H. & Yang, G.S. (2012) Transcriptome profile analysis of porcine adipose tissue by high-throughput sequencing. Animal Genetics 43, 144–152.

Liang, J., Fu, M., Ciociola, E., Chandalia, M. & Abate, N. (2007) Role of ENPP1 on Adipocyte Maturation. PLoS ONE 2, e882.

Litten-Brown, J.C., Corson, A.M. & Clarke, L. (2010) Porcine models for the metabolic syndrome, digestive and bone disorders: a general overview. Animal 4, 899–920.

Lumeng, C.N. & Saltiel, A.R. (2011) Inflammatory links between obesity and metabolic disease. J. Clin. Invest. 121, 2111–2117.

McCurdy, C.E. & Klemm, D.J. (2013) Adipose tissue insulin sensitivity and macrophage recruitment: Does PI3K pick the pathway? Adipocyte 2, 135–142.

853 Mehta, J.L. & Li, D. (2002) Identification, regulation and function of a novel lectin-like
854 oxidized low-density lipoprotein receptor. J. Am. Coll. Cardiol. 39, 1429–1435.
855 Meyre, D., Bouatia-Naji, N., Tounian, A., Samson, C., Lecoeur, C., Vatin, V., Ghoussaini, M.,
856 Wachter, C., Hercberg, S., Charpentier, G., Patsch, W., Pattou, F., Charles, M.-A.,
857 Tounian, P., Clément, K., Jouret, B., Weill, J., Maddux, B.A., Goldfine, I.D., Walley, A.,
858 Boutin, P., Dina, C. & Froguel, P. (2005) Variants of ENPP1 are associated with childhood
859 and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. Nat
860 Genet 37, 863–867.
861 Mikawa, A., Suzuki, H., Suzuki, K., Toki, D., Uenishi, H., Awata, T. & Hamasima, N. (2004)
862 Characterization of 298 ESTs from porcine back fat tissue and their assignment to the
863 SSRH radiation hybrid map. Mamm Genome 15, 315–322.
864 Moon, J.-K., Kim, K.-S., Kim, J.-J., Choi, B.-H., Cho, B.-W., Kim, T.-H. & Lee, C.-K. (2009)
865 Differentially expressed transcripts in adipose tissue between Korean native pig and
866 Yorkshire breeds. Anim. Genet. 40, 115–118.
867 Moreno-Navarrete, J.M., & Fernández-Real, J.M. (2012) Adipocyte Differentiation. In
868 Adipose Tissue Biology (ed. by M.E. Symonds). pp. 17-38. Springer, New York
869 Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R. & Schäffer, A.A. (2008)
870 Database indexing for production MegaBLAST searches. Bioinformatics 24, 1757–1764.
871 Newsholme, P. & de Bittencourt, P.I.H. (2014) The fat cell senescence hypothesis: a
872 mechanism responsible for abrogating the resolution of inflammation in chronic disease.
873 Current Opinion in Clinical Nutrition and Metabolic Care 17, 295–305.
874 Padilla, J., Jenkins, N.T., Lee, S., Zhang, H., Cui, J., Zuidema, M.Y., Zhang, C., Hill, M.A.,
875 Perfield, J.W., Ibdah, J.A., Booth, F.W., Davis, J.W., Laughlin, M.H. & Rector, R.S. (2013)

Vascular transcriptional alterations produced by juvenile obesity in Ossabaw swine. Physiol. Genomics 45, 434–446.

Pan, W., Ciociola, E., Saraf, M., Tumurbaatar, B., Tuvdendorj, D., Prasad, S., Chandalia, M. & Abate, N. (2011) Metabolic consequences of ENPP1 overexpression in adipose tissue. Am. J. Physiol. Endocrinol. Metab. 301, E901–911.

Peng, Y., Xiang, H., Chen, C., Zheng, R., Chai, J., Peng, J. & Jiang, S. (2013) MiR-224 impairs adipocyte early differentiation and regulates fatty acid metabolism. The International Journal of Biochemistry & Cell Biology 45, 1585–1593.

Quinlan, A.R. & Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. (2011a) Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27, 2325–2329.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L. & Pachter, L. (2011b) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12, R22.

Ropka-Molik, K., Zukowski, K., Eckert, R., Gurgul, A., Piórkowska, K. & Oczkowicz, M. (2014) Comprehensive analysis of the whole transcriptomes from two different pig breeds using RNA-Seq method. Anim. Genet. 45, 674–684.

Rosen, E.D. & MacDougald, O.A. (2006) Adipocyte differentiation from the inside out. Nat. Rev. Mol. Cell Biol. 7, 885–896.

Russell, T.D., Palmer, C.A., Orlicky, D.J., Bales, E.S., Chang, B.H.-J., Chan, L. & McManaman, J.L. (2008) Mammary glands of adipophilin-null mice produce an amino-terminally truncated form of adipophilin that mediates milk lipid droplet formation and secretion. J. Lipid Res. 49, 206–216.

Russo, V., Buttazzoni, L., Baiocco, C., Davoli, M.R., Nanni Costa, N.L., Schivazappa, O.C. & Virgili, P.C. (2000) Heritability of muscular cathepsin B activity in Italian Large White pigs. Journal of Animal Breeding and Genetics 117, 37–42.

Russo, V., Fontanesi, L., Scotti, E., Beretti, F., Davoli, R., Nanni Costa, L., Virgili, R. & Buttazzoni, L. (2008) Single nucleotide polymorphisms in several porcine cathepsin genes are associated with growth, carcass, and production traits in Italian Large White pigs. Journal of Animal Science 86, 3300–3314.

Serlachius, M. & Andersson, L.C. (2004) Upregulated expression of stanniocalcin-1 during adipogenesis. Exp. Cell Res. 296, 256–264.Sethi, J.K. (2010) Activatin' human adipose progenitors in obesity. Diabetes 59, 2354–2357.

Sethi, J.K. (2010) Activatin' Human Adipose Progenitors in Obesity. Diabetes 59, 2354–2357.

Smith, S.H., Wilson, A.D., Van Ettinger, I., MacIntyre, N., Archibald, A.L. & Ait-Ali, T. (2014) Down-regulation of mechanisms involved in cell transport and maintenance of mucosal integrity in pigs infected with Lawsonia intracellularis. Vet. Res. 45, 55.

Sodhi, S.S., Song, K.-D., Ghosh, M., Sharma, N., Lee, S.J., Kim, J.H., Kim, N., Mongre, R.K., Adhikari, P., Kim, J.Y., Hong, S.P., Oh, S.J. & Jeong, D.K. (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, 233–242.

Spurlock, M.E. & Gabler, N.K. (2008) The development of porcine models of obesity and the metabolic syndrome. J. Nutr. 138, 397–402.

Toedebusch, R.G., Roberts, M.D., Wells, K.D., Company, J.M., Kanosky, K.M., Padilla, J., Jenkins, N.T., Perfield, J.W., Ibdah, J.A., Booth, F.W. & Rector, R.S. (2014) Unique transcriptomic signature of omental adipose tissue in Ossabaw swine: a model of childhood obesity. Physiol. Genomics 46, 362–375.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. & Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562–578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. & Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515.

Trayhurn, P. (2005) Endocrine and signalling role of adipose tissue: new perspectives on fat. Acta Physiol. Scand. 184, 285–293.

Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H., Okumura, N., Hamasima, N. & Awata, T. (2004) PEDE (Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucl. Acids Res. 32, D484–D488.

Uenishi, H., Eguchi-Ogawa, T., Shinkai, H., Okumura, N., Suzuki, K., Toki, D., Hamasima, N. & Awata, T. (2007) PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10 147 porcine full-length cDNA sequences. Nucl. Acids Res. 35, D650–D653.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.

Varga, O., Harangi, M., Olsson, I. a. S. & Hansen, A.K. (2010) Contribution of animal models to the understanding of the metabolic syndrome: a systematic overview. Obes Rev 11, 792–807.

Vicario, M., Skaper, S.D., Negro, A. (2014) The small heat shock protein HspB8: role in nervous system physiology and pathology. CNS Neurol Disord Drug Targets 13, 885–895.

Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., Idowu, O., Li, M., Shen, C., Hu, A., Haydon, R.C., Kang, R., Mok, J., Lee, M.J., Luu, H.L. & Shi, L.L. (2014) Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes Dis 1, 87–105.

Yamada, K., Santo-Yamada, Y. & Wada, K. (2002) Restraint stress impaired maternal behavior in female mice lacking the neuromedin B receptor (NMB-R) gene. Neurosci. Lett. 330, 163–166.

Zhou, C., Zhang, J., Ma, J., Jiang, A., Tang, G., Mai, M., Zhu, L., Bai, L., Li, M. & Li, X. (2013) Gene expression profiling reveals distinct features of various porcine adipose tissues. Lipids Health Dis 12, 75.

Zhou, Z.-Y., Li, A.-M., Adeola, A.C., Liu, Y.-H., Irwin, D.M., Xie, H.-B. & Zhang, Y.-P. (2014) Genome-Wide Identification of Long Intergenic Noncoding RNA Genes and Their Potential Association with Domestication in Pigs. Genome Biol Evol 6, 1387–1392.

Supporting information

Additional supporting information may be found in the online version of this article.

Supplementary Tables are included in the file:

TranscriptomeILW_SupplementaryTables.xlsx

Table S1 - Primers and PCR condition used for the validation.

EXT: primer pairs used for the amplification of a larger PCR product

INT: primer pairs used for the creation of the standard curve and for the qRT-PCR analysis

Table S2 - Number of reads for each sample

For each sample is indicated the total raw reads sequenced, total clean reads after the trimming and length filters and total reads mapped to the reference genome. Reported values refer to reads as they were single end (total clean paired reads are half the value in the table). Respective percentages are shown in the last three columns.

Table S3 - Types of transcripts expressed in backfat tissue, according to the considered genome annotations.

Transcripts, associated to eight Cufflinks class codes (see http://cufflinks.cbcb.umd.edu/manual.html#class_codes), were classified into three major informative groups.

Table S4 – Intergenic transcript annotations.

Table S5 - Transcript coding potential predicted by Coding Potential Calculator (CPC) for intergenic transcripts.

Reliable noncoding: CPC score <-1

Noncoding: CPC score -1=> / <=0

Coding: CPC score >0

Table S6 - Most expressed transcripts (top 75%) detected in porcine backfat.

Table S7 - David functional annotation clustering of the most expressed genes.

The 10 most relevant clusters are reported

Table S8 - List of differentially expressed genes detected by Cuffdiff2.

Table S9 - List of differentially expressed genes detected by DeSeq2.

Table S10 - List of differentially expressed transcripts detected by Cuffdiff2.

Table S11 - Transcript coding potential predicted by Coding Potential Calculator (CPC) for the differentially expressed transcripts.

Reliable noncoding: CPC score <-1

Noncoding: CPC score -1=> / <=0

Coding: CPC score >0

Figure S1 - Read processing and alignment results.

File: TranscriptomeILW_FigureS1.jpg

(A) The boxplots show the distribution of the reads considered in different steps and filters of the computational analysis pipeline, in the 20 considered samples. From left to right we show the number of raw reads sequenced, of clean reads resulted from the filtering steps, of reads successfully mapped to the reference genome, and of reads with unique alignment in the genome. (B) From the left, the bars show the average amounts, in the 20 considered samples, of reads spliced, aligned to an exon, to an intron, to intergenic regions (according to the *Sus scrofa* 10.2 genome annotation), or spanning exon-intron borders. Different colors indicate the proportion of read aligning to chromosomes (blue), genome scaffolds (red) or mitochondrial genome (yellow). (C) Number of expressed genes detected in different chromosomes, in mitochondrial genome (Mt) or in genome scaffolds (S).

Figure S2 – Gene expression distribution in FAT and LEAN groups.

File: TranscriptomeILW_FigureS2.jpg

Cumulative gene expression is shown for the two groups. The figure represents the number of genes (horizontal axis) required to reach different percentages (vertical axis) of the overall gene expression. The inner panel focus on the cumulative expression curves for 50% and 75% of the expression.

Figure S3 – Alignment of the four detected isoforms of *PLIN2* gene (red box) with the porcine and vertebrates transcripts present in Ensembl.

File: TranscriptomeILW_FigureS3.jpg

Figure S4 – Principal component analysis (PCA) based on gene expression profiles.

File: TranscriptomeILW_FigureS4.jpg

The figure presents sample separation according to the two principal components, explaining most of the gene expression variation in the data. Samples are represented by dots, with green and orange colours indicating LEAN and FAT samples, respectively in Panel A) and red and blue indicating females and castrated males in Panel B). The PCA shows a clear separation of LEAN and FAT samples, with no separation of samples by sex.

Figure captions

Figure 1 – Transcripts and isoforms classification.

TranscriptomeILW_Figure1.jpg

(A) Expressed transcript were classified, according to current gene annotations, into 8 types, reported with different colors (see legend) and grouped in three categories: K (known) collects transcripts found in reference annotation (yellow); I (isoform) collects alternative forms of transcripts (red shades); N collects new transcripts from not-annotated loci (green shades). The pie chart shows the number of transcripts detected, for each type, and their mutual proportions. Three transcript types of the N group have few elements (43 intronic; 5 possible polymerase run-on fragments; 3 transcript intron overlap a reference intron on the opposite strand) and are barely visible in the chart. (B) Transcript length distributions in the three categories. (C) Transcript expression level distribution for the three categories. (D) Number of genes (vertical axis) with their number of transcript isoforms detected (horizontal axis). Genes with only one transcript isoforms detected are the most frequent; however, genes with up to 31 different isoforms were detected. (E) The proportion of each transcript type for the transcript isoforms grouped as in (D). Genes with only one isoform (first bar) are mainly intergenic genes (green part). For genes having more than one isoform expressed, the proportion of novel isoforms detected increases along with the number of different isoforms for a gene (red part).

Figure 2 - Coding potential of new intergenic transcripts.

TranscriptomeILW_Figure2.jpg

According to CPC scores, calculated both for the forward and for the reverse complement sequence, the intergenic transcripts were classified as “coding”, “non-coding” and “reliable non-coding”. (A) The pie chart shows numbers and proportions of intergenic transcripts falling in each category and provides the color code for the figure panels. (B) and (C) show respectively the distribution of lengths and of expression levels of intergenic transcripts, binned in the three categories. (D) Percentages of transcripts per category are compared, considering all the intergenic transcripts and the subset of the intergenic transcripts ranked within the 10% most expressed transcripts considering the whole transcriptome.

Figure 3 – Differentially expressed genes and transcripts identified.

TranscriptomeILW_Figure3.jpg

(A) Intersection of genes resulting differentially expressed (DE) according to DESeq2 and Cuffdiff2 analysis, and genes with at least one transcript resulting DE according to the transcript-level Cuffdiff2 analysis. We focused on the transcripts belonging to the 85 genes commonly identified by all the methods. (B) Proportions of the new and known DETs resulting higher- and lower-expressed in FAT *vs*. LEAN samples. (C) Number of DE genes mapping to chromosomes or to genome scaffolds (S).

Figure 4 - qRT-PCR validation of eleven genes differentially expressed according to RNA-seq data.

TranscriptomeILW_Figure4.jpg

(A) Log_2 FC values obtained from RNA-seq, according to Cuffdiff2 estimates, (black bars) and from qRT-PCR data (grey bars), for the eleven tested genes; (B) scatterplot showing the good correlation between the Log_2 FC values calculated with the two experimental methods.

Figure 5 – Genes differentially expressed between FAT and LEAN animals impact on specific and connected biological processes.

TranscriptomeILW_Figure5.jpg

Genes differentially expressed in FAT *vs*. LEAN pigs converge to specific functions that are more activated or impaired in FAT pigs. Genes and functions upregulated and downregulated in FAT pigs are shown in red and green shades, respectively. Several genes more expressed in FAT pigs are linked to fat deposition and lipid metabolism, to adipocyte differentiation and maturation or to signaling pathways regulating them; FAT pigs show as well increased expression of genes involved in inflammation and immunity and increased expression of genes involved in the control of complex behavior, also by inflammation-mediated secretory activity of adipocytes. Metabolic alterations induce chronic stress in the adipose tissue. FAT pigs shows under-expression of several genes involved in stress response by unfolded protein binding and misfolded protein aggregation prevention. The

impairment of these functions might in turn augment inflammation and the consequent secretory activity and possibly induce senescence.

Tables

Table 1 - Genetic indexes and phenotypes for BFT and hot carcass weight of the pigs selected for the transcriptome analysis.

Group	Sample ID	Sex	Day of slaughter	Slaughter weight (kg) (*)	BFT phenotype (mm)	BFT EBV		
							Mean	SD
FAT	477	M	6	120	43	7.36	5.22	1.3
	476	F	6	119	37	7.17		
	474	M	2	113	38	6.03		
	482	F	9		42	5.75		
	478	F	7	118	33	5.05		
	516	F	3	115	36	4.88		
	479	M	8		41	4.76		
	483	F	10	119	38	4.41		
	489	M	18	108	35	3.54		
	484	M	15	128	35	3.27		
LEAN	490	M	19	113	24	-6.46	-8.63	1.4
	473	F	2	132	23	-7.54		
	487	M	18	110	23	-7.61		
	517	M	4	117	20	-7.71		
	485	F	17	126	20	-7.82		
	475	M	5	119	20	-8.03		
	481	M	9		22	-9.91		
	486	F	17	123	19	-10.27		
	488	F	18	128	19	-10.37		
	480	F	9		16	-10.59		

EBV: estimated breeding value

BFT: backfat thickness.

(*) slaughter weight: the hot carcass slaughter weight is reported. For four animals the weight is not available due to a problem of the automatic recording system at the slaughterhouse.

Table 2 - List of the DE genes and transcripts.

Cufflinks transcript ID	Cufflinks gene ID	Gene locus	Gene symbol	Cuffdiff2 gene log2 FC FAT *vs.* LEAN	Transcript group	Coding potential
TCONS_00102010	XLOC_040987	JH118612.1:113132-140205	*DSC2*	2.55	Known	-
TCONS_00061823	XLOC_023331	4:78928264-78930654	-	2.46	New	NON CODING
TCONS_00033774	XLOC_013001	15:140797584-140847461	*NYAP2*	2.38	New	CODING
TCONS_00061359	XLOC_023211	4:35670339-35685878	*DCSTAMP*	2.23	Novel isoform	CODING
TCONS_00095554	XLOC_036823	GL893451.1:11131-27485	*CRLF2*	2.21	Known	-
TCONS_00093244	XLOC_035190	9:50996895-51001264	-	2.17	New	NON CODING
TCONS_00087029	XLOC_032796	8:140307937-140315415	*SPP1*	2.09	Known	-
TCONS_00003007	XLOC_000806	1:283547172-283552108	-	2.07	New	CODING
TCONS_00095549	XLOC_036822	GL893451.1:7060-10625	-	2.03	New	NON CODING
TCONS_00067029	XLOC_025404	5:36179189-36186325	*LYZ*	2.03	Known	-
TCONS_00042581	XLOC_016514	18:6731368-6733669	*GIMAP2*	1.98	Known	-
TCONS_00061600	XLOC_023265	4:55660234-55715444	*ATP6V0D2*	1.96	Novel isoform	CODING
TCONS_00039556	XLOC_015432	17:53815353-53827092	*MMP9*	1.92	Known	-
TCONS_00039900	XLOC_015518	17:4110395-4192029	*MSR1*	1.92	Known	-
TCONS_00061643	XLOC_023283	4:62172539-62226917	*STMN2*	1.85	Known	-
TCONS_00034645	XLOC_013236	15:62409564-62414328	-	1.84	New	RELIABLE NON CODING
TCONS_00091509	XLOC_034399	9:63158999-63198155	*ST14*	1.79	Novel isoform	CODING
TCONS_00098750	XLOC_038994	GL895411.1:0-1073	*INHBB*	1.65	New	CODING
TCONS_00022322	XLOC_008474	13:32323641-32330286	*CCR1*	1.63	Known	-
TCONS_00044383	XLOC_017319	2:11807281-11850646	*MPEG1*	1.63	Known	-
TCONS_00075056	XLOC_028007	6:70039585-70099223	*PADI2*	1.6	Known	-

TCONS_00095875	XLOC_037025	GL893645.1:0-307	-	1.57	New	RELIABLE NON CODING
TCONS_00084869	XLOC_032187	8:71288921-71302169	*AMBN*	1.56	Known	-
TCONS_00033691	XLOC_012975	15:133452328-133456736	*SLC11A1*	1.56	Known	-
TCONS_00089513	XLOC_033895	9:90266412-90348498	*SCIN*	1.55	Known	-
TCONS_00042660	XLOC_016535	18:8306789-8313120	-	1.52	New	CODING
TCONS_00059834	XLOC_022860	4:99905518-99915176	*CD1A*	1.52	Novel isoform	CODING
TCONS_00059837	XLOC_022860	4:99905518-99915176	*CD1A*	1.52	Known	-
TCONS_00093519	XLOC_035465	9:101443296-101443885	*GPNMB*	1.46	New	NON CODING
TCONS_00098157	XLOC_038614	GL894967.1:126-517	*GPNMB*	1.42	New	CODING
TCONS_00018804	XLOC_007247	12:23439824-23441829	-	1.4	New	CODING
TCONS_00103084	XLOC_041497	X:37303173-37393818	*CYBB*	1.38	Known	-
TCONS_00065337	XLOC_024931	5:52504178-52625145	*BCAT1*	1.37	Novel isoform	CODING
TCONS_00098113	XLOC_038589	GL894923.1:47-563	*GPNMB*	1.36	New	CODING
TCONS_00002441	XLOC_000664	1:227333991-227356844	*PLIN2*	1.32	Novel isoform	CODING
TCONS_00044392	XLOC_017322	2:12191483-12243400	*LPXN*	1.31	Known	-
TCONS_00084565	XLOC_032101	8:33970571-33982450	*UCHL1*	1.27	Novel isoform	CODING
TCONS_00067389	XLOC_025495	5:64579162-64590512	*OLR1*	1.26	Known	-
TCONS_00059747	XLOC_022835	4:97720982-97736619	*CD48*	1.25	Known	-
TCONS_00028769	XLOC_011055	14:143745489-143752509	*GMFG*	1.23	Known	-
TCONS_00029056	XLOC_011139	14:8804077-8816800	*STC1*	1.23	Novel isoform	CODING
TCONS_00098643	XLOC_038938	GL895339.1:13269-61205	*COTL1*	1.15	Known	-
TCONS_00100592	XLOC_040068	GL896326.1:1999-3913	*ACP5*	1.13	Known	-
TCONS_00096837	XLOC_037668	GL894123.1:0-400	*CD163*	1.13	New	CODING
TCONS_00097297	XLOC_037990	GL894401.1:0-471	*CD163*	1.13	New	CODING
TCONS_00005002	XLOC_001331	1:125897935-125953413	*AQP9*	1.09	Known	-
TCONS_00096863	XLOC_037686	GL894145.1:0-401	*CD163*	1.09	New	CODING
TCONS_00071337	XLOC_027094	6:74616232-74621248	*C1QC*	1.08	Known	-

TCONS_00012469	XLOC_005058	11:21534980-21685851	*LCP1*	1.07	Novel isoform	CODING
TCONS_00079920	XLOC_030238	7:94900207-94906867	*AKAP5, LOC100153460*	1.06	Novel isoform	CODING
TCONS_00041537	XLOC_016257	18:6613761-6621027	*GIMAP4*	1.06	Known	-
TCONS_00097908	XLOC_038444	GL894747.1:3047-10617	*HMOX1*	1.06	Novel isoform	CODING
TCONS_00030401	XLOC_011444	14:71516962-71521335	*EGR2*	1.05	Known	-
TCONS_00030878	XLOC_011579	14:117265093-117349965	*BLNK*	1.04	Known	-
TCONS_00056578	XLOC_021190	3:77408776-77439119	*PLEK*	1.04	Known	-
TCONS_00071335	XLOC_027093	6:74609911-74612993	*C1QA*	1.02	Known	-
TCONS_00081915	XLOC_030757	7:54395230-54406136	*BCL2A1*	1.01	Known	-
TCONS_00041554	XLOC_016261	18:6872940-6875292	*GIMAP1*	1	Known	-
TCONS_00085005	XLOC_032236	8:79743274-79751980	*SFRP2*	0.99	Known	-
TCONS_00098919	XLOC_039115	GL895590.1:0-1327	*GPNMB*	0.91	New	NON CODING
TCONS_00068526	XLOC_026077	5:52625315-52630242	*BCAT1*	0.89	New	RELIABLE NON CODING
TCONS_00062055	XLOC_023401	4:97099149-97103132	*FCER1G*	0.87	Known	-
TCONS_00009719	XLOC_003695	10:48841010-48961015	*MRC1*	0.86	Novel isoform	CODING
TCONS_00030894	XLOC_011584	14:117670639-117938624	*PIK3AP1*	0.85	Known	-
TCONS_00017526	XLOC_006800	12:36561025-36604089	*CLTC*	0.8	Novel isoform	CODING
TCONS_00062959	XLOC_023614	4:119674090-119703427	*CD53*	0.78	Known	-
TCONS_00081898	XLOC_030753	7:53623061-53644262	*CTSH*	0.78	Known	-
TCONS_00060570	XLOC_023035	4:119013307-119039899	*ADORA3*	0.74	Known	-
TCONS_00052401	XLOC_020144	3:11035819-11055510	*LAT2*	0.71	Known	-
TCONS_00004118	XLOC_001095	1:35133812-35137388	*CTGF*	0.68	Known	-
TCONS_00045043	XLOC_017499	2:59214054-59218018	*IFI30*	0.65	Known	-
TCONS_00004124	XLOC_001096	1:35240242-35281384	*ENPP1*	0.62	Known	-
TCONS_00062884	XLOC_023592	4:116704501-116707235	*OLFML3*	-0.54	Known	-
TCONS_00035484	XLOC_013426	15:131680309-131684630	*IGFBP5*	-0.65	Known	-

TCONS_00101718	XLOC_040809	JH118426.1:306724-312138	-	-0.77	New	RELIABLE NON CODING
TCONS_00063805	XLOC_024145	4:77261119-77264781	-	-0.77	New	NON CODING
TCONS_00050164	XLOC_018733	2:124815021-124828122	*CDO1*	-0.9	Novel isoform	CODING
TCONS_00101559	XLOC_040715	GL896532.1:212-2567	*ADSSL1*	-1.02	New	NON CODING
TCONS_00079927	XLOC_030240	7:94987617-94990126	*HSPA2*	-1.1	Known	-
TCONS_00083805	XLOC_031620	7:66542203-66555641	-	-1.18	New	CODING
TCONS_00041725	XLOC_016313	18:15292592-15295178	-	-1.61	New	CODING
TCONS_00048853	XLOC_018425	2:65175406-65180520	*DNAJB1*	-1.66	Novel isoform	CODING
TCONS_00029533	XLOC_011248	14:35688332-35701411	*HSPB8*	-1.81	Known	-
TCONS_00094194	XLOC_036009	GL892492.1:0-3540	*HSPA1B*	-2.32	New	NON CODING
TCONS_00101505	XLOC_040677	GL896522.1:9039-10877	*HSPA1A*	-2.57	New	RELIABLE NON CODING
TCONS_00098059	XLOC_038555	GL894890.1:5-696	*HSP70*	-3.44	New	NON CODING

Table 3 - David functional annotation clustering obtained considering the significant Biological Processes GO terms (Benjamini adjusted P-values <0.05) of genes more expressed in FAT than in LEAN animals.

Annotation Cluster 1	Enrichment Score: 7.0	
Term	Count	Genes
GO:0006954~inflammatory response	12	*C1QA, SLC11A1, CYBB, ADORA3, OLR1, HMOX1, CCR1, LYZ, C1QC, BLNK, CD163, SPP1*
GO:0006952~defense response	15	*ADORA3, OLR1, CCR1, LYZ, COTL1, C1QC, CD163, INHBB, CD48, C1QA, SLC11A1, CYBB, HMOX1, SPP1, BLNK*
GO:0009611~response to wounding	14	*ADORA3, PLEK, OLR1, CCR1, LYZ, C1QC, CD163, C1QA, SLC11A1, CYBB, CTGF, HMOX1, SPP1, BLNK*
GO:0009605~response to external stimulus	17	*ADORA3, PLEK, OLR1, CCR1, LYZ, C1QC, CD163, INHBB, C1QA, SLC11A1, CYBB, CTGF, SFRP2, HMOX1, STC1, SPP1, BLNK*
GO:0050896~response to stimulus	29	*ADORA3, AQP9, ENPP1, CCR1, UCHL1, ACP5, C1QC, CD48, SLC11A1, PLIN2, CTGF, HMOX1, FCER1G, BLNK, SPP1, EGR2, OLR1, PLEK, LYZ, CD1A, COTL1, CD163, INHBB, C1QA, CYBB, LAT2, SFRP2, STC1, LCP1*
GO:0006950~response to stress	19	*ADORA3, AQP9, PLEK, OLR1, CCR1, UCHL1, LYZ, COTL1, C1QC, CD163, INHBB, CD48, C1QA, SLC11A1, CYBB, CTGF, HMOX1, SPP1, BLNK*
Annotation Cluster 2	Enrichment Score: 2.7	
Term	Count	Genes
GO:0001775~cell activation	7	*CD48, SLC11A1, LAT2, PLEK, LCP1, BLNK, GIMAP1*
GO:0002274~myeloid leukocyte activation	4	*CD48, SLC11A1, LAT2, GIMAP1*
GO:0046649~lymphocyte activation	6	*CD48, SLC11A1, LAT2, LCP1, BLNK, GIMAP1*
Annotation Cluster 3	Enrichment Score: 2.4	

Term	Count	Genes
GO:0048583~regulation of response to stimulus	10	*C1QA, SLC11A1, LAT2, PLEK, ENPP1, HMOX1, FCER1G, C1QC, SPP1, GIMAP1*
GO:0050776~regulation of immune response	7	*C1QA, SLC11A1, LAT2, HMOX1, FCER1G, C1QC, GIMAP1*
GO:0050778~positive regulation of immune response	6	*C1QA, SLC11A1, LAT2, FCER1G, C1QC, GIMAP1*
GO:0002443~leukocyte mediated immunity	5	*C1QA, SLC11A1, LAT2, FCER1G, C1QC*
GO:0002682~regulation of immune system process	8	*C1QA, SLC11A1, LAT2, HMOX1, SCIN, FCER1G, C1QC, GIMAP1*
Annotation Cluster 4	Enrichment Score: 2.0	
Term	Count	Genes
GO:0060348~bone development	6	*AMBN, CTGF, ACP5, STC1, GPNMB, SPP1*
GO:0031214~biomineral formation	4	*AMBN, ENPP1, GPNMB, SPP1*
GO:0001503~ossification	5	*AMBN, CTGF, STC1, GPNMB, SPP1*
GO:0001501~skeletal system development	7	*AMBN, CTGF, MMP9, ACP5, STC1, GPNMB, SPP1*
Annotation Cluster 5	Enrichment Score: 1.6	
Term	Count	Genes
GO:0001775~cell activation	7	*CD48, SLC11A1, LAT2, PLEK, LCP1, BLNK, GIMAP1*

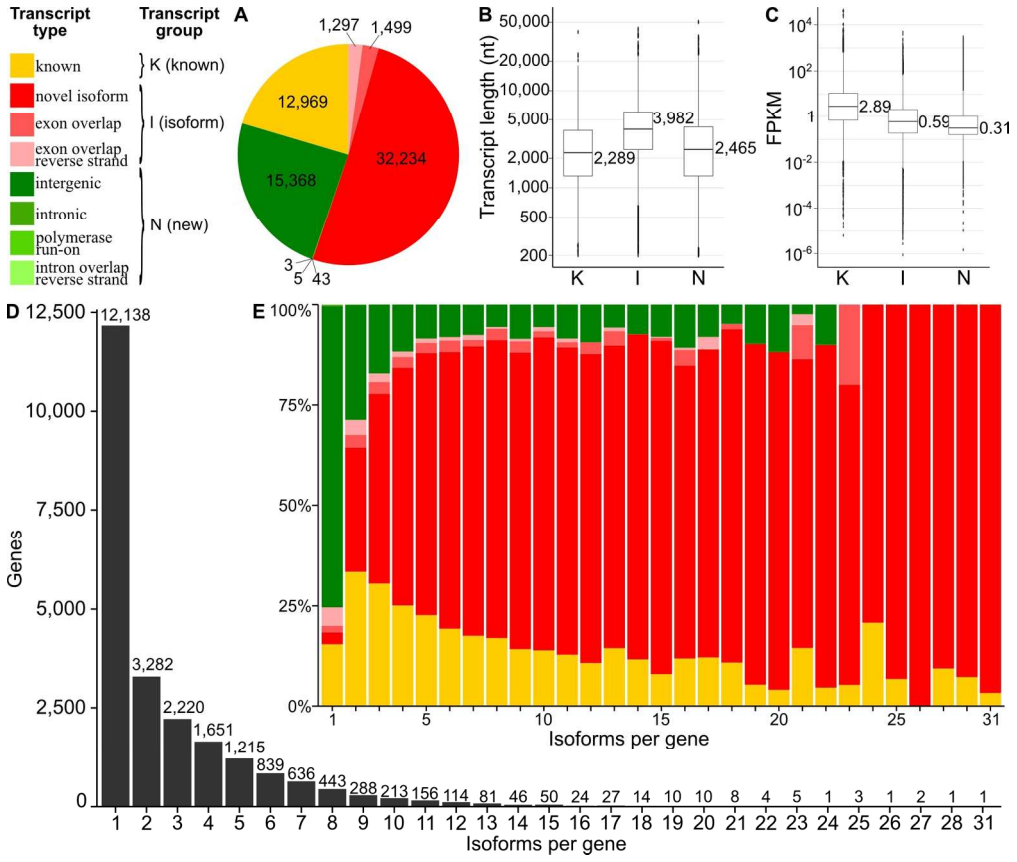

Transcript type
Transcript group
known
novel isoform
exon overlap
exon overlap reverse strand
intergenic
intronic
polymerase run-on
intron overlap reverse strand
K (known)
I (isoform)
N (new)
A
1,297
1,499
12,969
32,234
15,368
3
5
43
B
Transcript length (nt)
50,000
20,000
10,000
5,000
2,000
1,000
500
200
2,289
3,982
2,465
K
I
N
C
FPKM
10^4
10^2
1
10^-2
10^-4
10^-6
2.89
0.59
0.31
K
I
N
D
Genes
12,500
10,000
7,500
5,000
2,500
0
12,138
3,282
2,220
1,651
1,215
839
636
443
288
213
156
114
81
46
50
24
27
14
10
10
8
4
5
1
3
1
2
1
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 31
Isoforms per gene
E
100%
75%
50%
25%
0%
1 5 10 15 20 25 31
Isoforms per gene

Figure 1

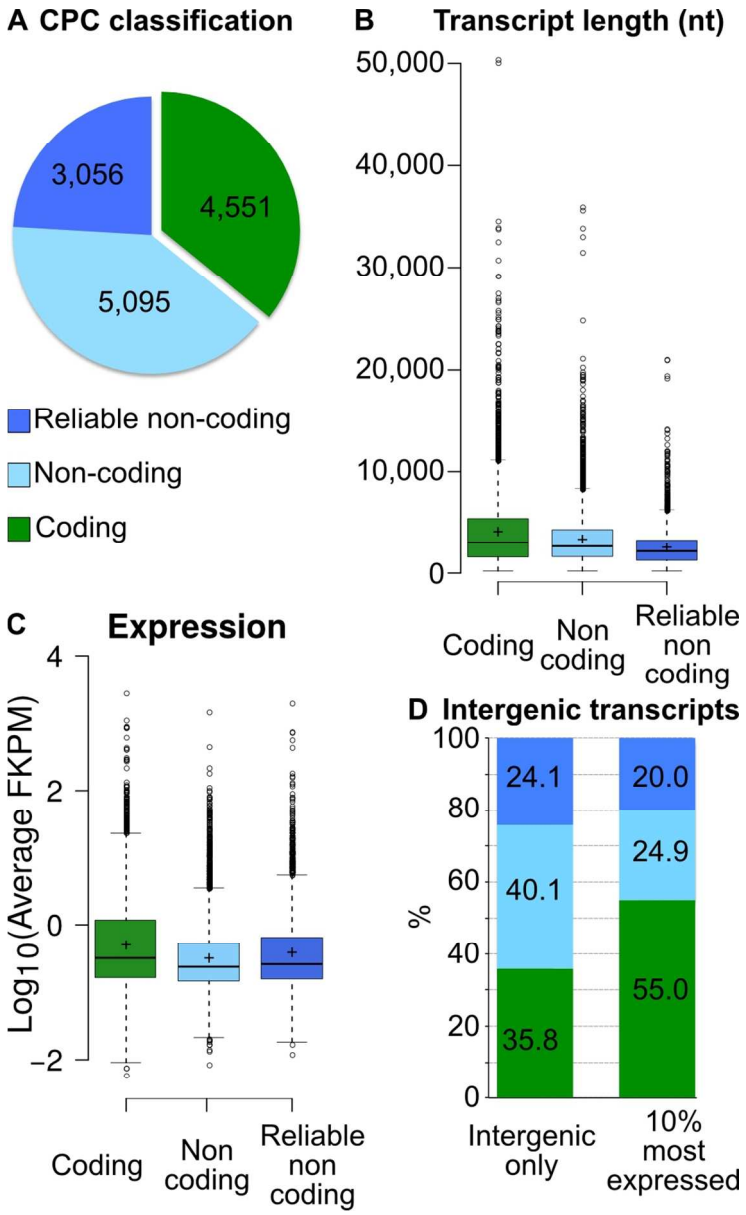
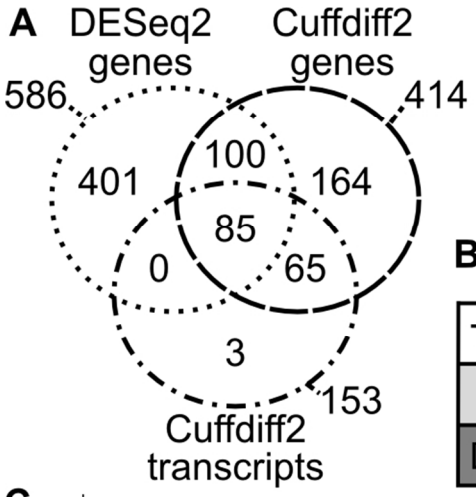

A CPC classification
3,056
4,551
5,095
Reliable non-coding
Non-coding
Coding
B Transcript length (nt)
50,000
40,000
30,000
20,000
10,000
0
Coding
Non coding
Reliable non coding
C Expression
Log10(Average FKPM)
4
2
0
−2
Coding
Non coding
Reliable non coding
D Intergenic transcripts
%
100
80
60
40
20
0
24.1
40.1
35.8
20.0
24.9
55.0
Intergenic only
10% most expressed

Figure 2



B

	DETs	Known	New
TOTAL	86	60	26
UP	72	54	18
DOWN	14	6	8

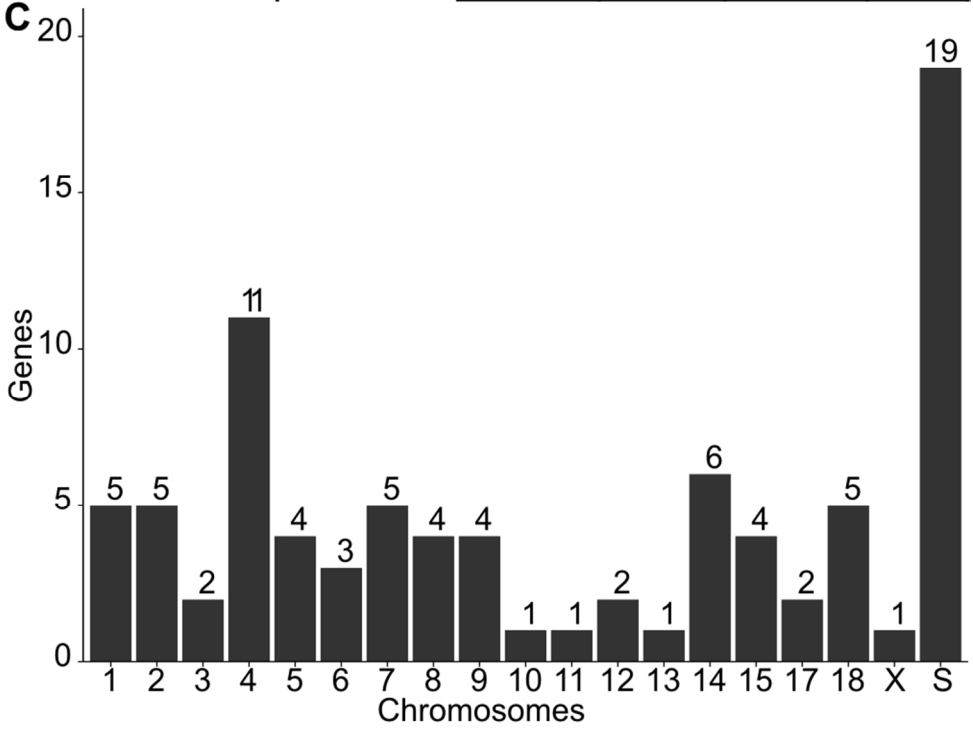


Figure 3
85x105mm (300 x 300 DPI)

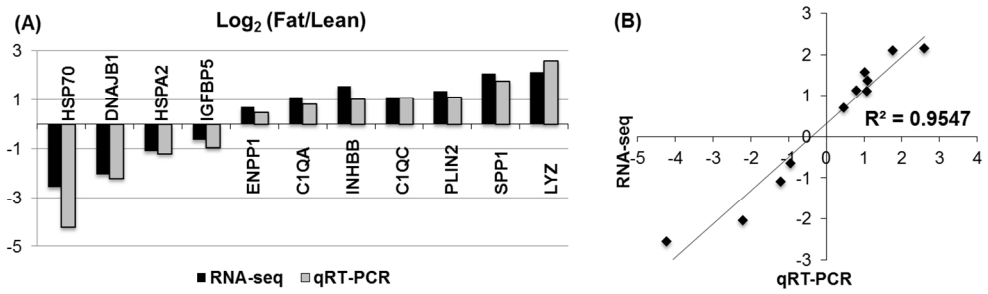


Figure 4
242x73mm (150 x 150 DPI)

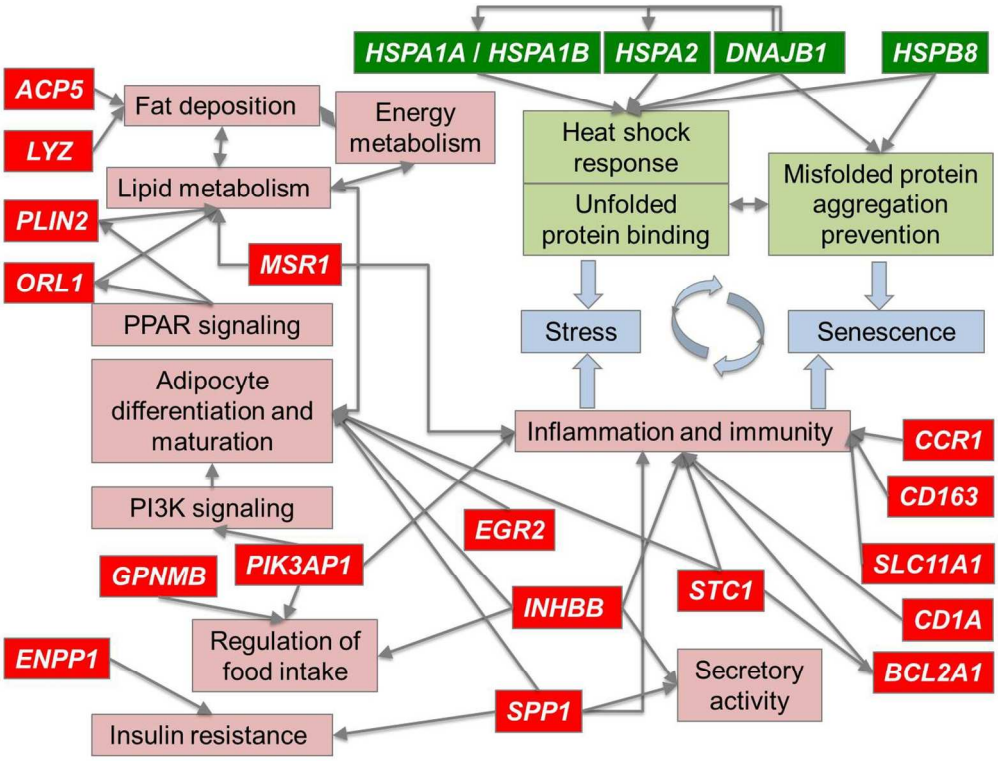


Figure 5
249x190mm (150 x 150 DPI)