

RESEARCH ARTICLE

Open Access



Differential expression of mRNA isoforms in the skeletal muscle of pigs with distinct growth and fatness profiles

Tainã Figueiredo Cardoso^{1,2}, Raquel Quintanilla³, Anna Castelló^{1,4}, Rayner González-Prendes¹, Marcel Amills^{1,4*} and Ángela Cánovas^{5*}

Abstract

Background: The identification of genes differentially expressed in the skeletal muscle of pigs displaying distinct growth and fatness profiles might contribute to identify the genetic factors that influence the phenotypic variation of such traits. So far, the majority of porcine transcriptomic studies have investigated differences in gene expression at a global scale rather than at the mRNA isoform level. In the current work, we have investigated the differential expression of mRNA isoforms in the *gluteus medius* (GM) muscle of 52 Duroc HIGH (increased backfat thickness, intramuscular fat and saturated and monounsaturated fatty acids contents) and LOW pigs (opposite phenotype, with an increased polyunsaturated fatty acids content).

Results: Our analysis revealed that 10.9% of genes expressed in the GM muscle generate alternative mRNA isoforms, with an average of 2.9 transcripts per gene. By using two different pipelines, one based on the CLC Genomics Workbench and another one on the STAR, RSEM and DESeq2 softwares, we have identified 10 mRNA isoforms that both pipelines categorize as differentially expressed in HIGH vs LOW pigs (P -value < 0.01 and ± 0.6 log₂fold-change). Only five mRNA isoforms, produced by the *ITGA5*, *SEMA4D*, *LITAF*, *TIMP1* and *ANXA2* genes, remain significant after correction for multiple testing (q -value < 0.05 and ± 0.6 log₂fold-change), being upregulated in HIGH pigs.

Conclusions: The increased levels of specific *ITGA5*, *LITAF*, *TIMP1* and *ANXA2* mRNA isoforms in HIGH pigs is consistent with reports indicating that the overexpression of these four genes is associated with obesity and metabolic disorders in humans. A broader knowledge about the functional attributes of these mRNA variants would be fundamental to elucidate the consequences of transcript diversity on the determinism of porcine phenotypes of economic interest.

Keywords: Alternative splicing, mRNA isoform, Swine, Differential expression

Background

Recent estimates indicate that in mammals, at least 70% of genes have multiple polyadenylation sites, > 50% of genes have alternative transcription start sites and nearly 95% of genes undergo alternative splicing (AS) yielding multiple messenger ribonucleic acid (mRNA) isoforms

[1, 2]. The use of alternative transcriptional initiation and/or termination sites can produce diverse pre-mRNAs, which can further be subjected to AS yielding a broad array of mRNA isoforms that are derived from a single gene. A recent study indicated that alternative transcription start and termination sites, rather than AS, encompasses most of tissue-dependent exon usage [1]. Transcripts produced by any of the mechanisms mentioned above might contribute to differences between tissues or cells by modifying protein structure and expression [3–5]. Indeed, the differential expression of mRNA isoforms has been associated with a broad array

* Correspondence: marcel.amills@uab.cat; acanovas@uoguelph.ca

¹Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

⁵Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada

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



of physiological and pathological conditions in humans [3, 4] and domestic species [5, 6].

The important consequences of transcript diversity on porcine phenotypes of economic interest have been recently evidenced in a couple of studies. In White Duroc × Erhualian F₂ intercross pigs, a mutation in a splice acceptor site of intron 9 (g.8283C > A) of the porcine phosphorylase kinase catalytic subunit gamma 1 (*PHKG1*) gene has been shown to drive the synthesis of an aberrant transcript subjected to nonsense-mediated decay [7]. This results in the inactivation of this enzyme, which plays a key role in the degradation of glycogen, and in the production of a low quality meat with a poor water-holding capacity [7]. Moreover, Koltes et al. [8], identified a mutation located in the pig guanylate binding protein 5 (*GBP5*) gene that introduces a new splice acceptor site that results in the insertion of five additional nucleotides, thus altering the open reading frame and introducing a premature stop-codon. This mutation has a major effect on the host response to the porcine respiratory and reproductive syndrome virus [8].

Transcript diversity of the porcine muscle has been poorly characterized so far and the majority of studies comparing the transcriptomes of pigs with distinct phenotypic attributes have just focused on global differences in gene expression, rather than identifying the specific transcripts that are differentially expressed (DE) [9–12]. The goals of the current experiment were to provide a first picture of transcript diversity in the *gluteus medius* (GM) muscle of pigs as well as to identify mRNA isoforms that are DE in the GM muscle of Duroc swine with distinct growth and fatness profiles.

Methods

Animal material

The muscle transcriptomes of 56 Duroc pigs, retrieved from a population of 350 individuals distributed in 5 half-sib families, were analyzed using RNA-sequencing (RNA-seq) (Additional file 1: Table S1). As previously reported by Gallardo et al. [13], barrows were transferred to the IRTA-CCP experimental test station after weaning (3–4 weeks of age) and bred under normal intensive conditions. In the first stage of fattening (up to 90 kg of live weight, around 150 days of age) barrows were fed ad libitum a standard diet with 18% protein, 3.8% fiber, 7.0% fat, 1.0% lysine, and 0.3% methionine (net energy concentration: 2450 kcal/kg). In the last period of fattening (i.e. 30–40 days before slaughter) animals were fed ad libitum a standard diet with 15.9% protein, 4.5% fiber, 5.2% fat, 0.7% lysine, and 0.2% methionine (net energy concentration: 2375 kcal/kg). Pigs were slaughtered when they reached ≈ 122 kg live weight (i.e. at an age of 180–200 days approximately). Backfat and ham fat

thickness were measured with a ruler in the cutting room 24 h after slaughtering. Lean meat content was estimated on the basis of fat and muscle thickness data measured with an Autofom ultrasound device. Samples of the GM muscle were retrieved, snap frozen in liquid nitrogen and stored at –80 °C. A near infrared transmittance device (NIT, Infratec 1625, Tecator Hoganas, Sweden) was employed to determine intramuscular fat content. The determination of fatty acid composition was achieved with a technique based on the gas chromatography of methyl esters [14]. As reported by Gallardo and coworkers [13], blood samples were obtained at 190 days and a variety of enzymatic methods were used to determine cholesterol (cholesterol oxidase-based method), high-density lipoprotein (immunoinhibition method) and triglyceride concentrations (glycerol kinase reaction). Low density lipoprotein concentration was calculated according to the equation of Friedewald et al. [15].

Principal component analysis based on the 13 traits listed in Table 1 was performed in order to select pigs with distinct growth and fatness phenotypes (HIGH and LOW pigs) [10]. When compared with LOW pigs, the HIGH ($n = 28$) ones showed a higher live weight, backfat thickness and intramuscular fat content and also displayed increased serum lipid concentrations and muscle saturated (SFA) and monounsaturated (MUFA) fatty acids contents (Table 1). On the other hand, LOW pigs ($n = 28$), were lighter, leaner and had a higher muscle polyunsaturated fatty acids (PUFA) content than HIGH pigs.

RNA isolation, library construction and sequencing

Each muscle sample ($N = 56$, 28 HIGH and 28 LOW) was individually submerged in liquid nitrogen and grinded with a mortar and a pestle to produce a homogenous powder. This powder was submerged in TRIzol reagent (Thermo Fisher Scientific, Barcelona, Spain) and homogenized with a Polytron device (IKA, Staufen, Germany). Total RNA was purified with the Ambion RiboPure kit (Thermo Fisher Scientific, Barcelona, Spain) by following the instructions of the manufacturer. RNA samples were resuspended in a buffer solution provided in the kit and kept at –80 °C until use. RNA quantification and purity were assessed with a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Barcelona, Spain), while integrity was checked with a Bioanalyzer-2100 equipment (Agilent Technologies, Santa Clara, CA). All samples showed an RNA integrity number above 7.5. Sequencing libraries were prepared with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) and sequenced in a paired-end mode (2×75 bp), multiplexing two samples in each sequencing lane, on a HiSeq2000 Sequencing System

Table 1 Mean values \pm standard deviation (SD) for 13 phenotypes recorded in HIGH and LOW Duroc pigs

Phenotypes	HIGH group (N = 28)	LOW group (N = 28)
	Mean \pm SD	Mean \pm SD
Carcass traits		
LW - Live weight (kg)	130.90 \pm 9.46 ^a	110.75 \pm 16.62 ^b
BFTiv - Backfat thickness in vivo (mm)	28.74 \pm 3.47 ^a	18.76 \pm 3.90 ^b
BFT - Backfat thickness 3rd-4th ribs (mm)	47.07 \pm 11.94 ^a	33.89 \pm 10.03 ^b
HFT - Ham fat thickness (mm)	28.02 \pm 2.70 ^a	20.97 \pm 3.56 ^b
LEAN - Lean content (%)	39.17 \pm 5.15 ^a	45.48 \pm 4.21 ^b
Meat quality traits (<i>gluteus medius</i>)		
IMF - Intramuscular fat content (%)	7.27 \pm 1.70 ^a	3.69 \pm 0.93 ^b
SFA - Saturated fatty acids content (%)	38.70 \pm 1.41 ^a	34.76 \pm 1.30 ^b
PUFA - Polyunsaturated fatty acids content (%)	14.71 \pm 3.08 ^a	27.82 \pm 4.40 ^b
MUFA - Monounsaturated fatty acids content (%)	46.58 \pm 2.67 ^a	37.4 \pm 4.30 ^b
Serum lipid levels - 190 days		
CHOL - Total cholesterol (mg/dL)	161.11 \pm 30.32 ^a	104.17 \pm 16.40 ^b
HDL - HDL-cholesterol (mg/dL)	61.12 \pm 8.58 ^a	42.92 \pm 9.19 ^b
LDL - LDL-cholesterol (mg/dL)	86.34 \pm 29.32 ^a	50.57 \pm 15.12 ^b
TG - Triacylglycerides (mg/dL)	68.07 \pm 26.28 ^a	50.71 \pm 29.70 ^b

Means with different letters are significantly different (P -value $<$ 0.05), t-test for: LW, IMF, MUFA, CHOL and LDL; Wilcoxon test for: BFTiv, BFT, LEAN, SFA, PUFA, HDL and TG

(Illumina, San Diego, CA). Library preparation and sequencing were developed according to the protocols recommended by the manufacturer.

Differential expression analyses of mRNA isoforms between HIGH and LOW pigs

Adaptors and low quality bases were trimmed from sequences by using Trimmomatic [16] with default parameters. Quality control of sequences in FASTQ and BAM format was assessed with the FASTQC software (Babraham Bioinformatics, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sequence quality was measured by taking into account sequence-read lengths and base-coverage (distribution = 75 bp, 100% coverage in all bases), nucleotide contributions and base ambiguities (GC-content \sim 50%, \sim 25% of A, T, G and C nucleotide contributions and an ambiguous base-content $<$ 0.1%) and a Phred score higher than 30 (i.e. base-calling accuracy larger than 99.9%). All samples, except four, passed the quality control parameters, so our final data set consisted of 52 animals. With the aim of minimizing the rate of false positives, we used two different pipelines in the analysis of differential expression. In the first pipeline,

read mapping and counting were carried out with CLC Genomics Workbench 8.5 (CLC Bio, Aarhus, Denmark, <https://www.qiagenbioinformatics.com/>). In the second pipeline, reads were mapped with Spliced Transcripts Alignment to a Reference (STAR) v. 2.4 [17], counted with the RNA-Seq by Expectation Maximization (RSEM) software v. 1.3 [18] and differential expression was analysed with DESeq2 [19]. We considered as DE mRNA isoforms those simultaneously identified with the two pipelines.

Pipeline 1 (CLC genomics workbench)

The Large Gap Mapper (LGM) tool of CLC Genomics Workbench 8.5 was used to map the reads. This tool can map sequence reads that span introns without requiring prior transcript annotations. In this way, the LGM tool finds the best match for a given read. If there is an unaligned end which is long enough for the mapper to handle (17 bp for standard mapping) this segment of the read is re-mapped with the standard read mapper of the CLC Genomics Workbench. This process is repeated until no reads have unaligned ends that are longer than 17/18 bp. In our study, short sequence reads were mapped and annotated by using as template the pig reference genome version 10.2 (*Sscrofa* 10.2 - <http://www.ensembl.org/info/data/ftp/index.html>). Additional details can be found in http://resources.qiagenbioinformatics.com/manuals/transcriptdiscovery/208/index.php?manual=Large_gap_mapper.html. For mapping purposes, we considered alignments with a length fraction of 0.7 and a similarity fraction of 0.8. Two mismatches and three insertions and deletions per read were allowed. The quantification of mRNA isoform levels by the CLC Genomics Workbench follows a count-based model, where reads are counted on small counting units (exons), instead of the whole transcript unit, and the two possible splicing outcomes (inclusion and/or exclusion) are tested for each counting unit. Normalized count values are transformed on a decimal logarithmic scale. Statistical analysis of differential expression of splicing variants is based on an empirical analysis of digital gene expression [20], that implements an 'Exact Test' for two-group comparisons, assuming a negative binomial distribution and an overdispersion caused by biological variability estimated at 5%.

Pipeline 2 (STAR/RSEM/DESeq2)

The STAR software v. 2.4 [17] was employed to map the reads generated in the RNA-Seq experiment. The STAR algorithm comprises two main steps. First, a sequential maximum mappable seed search is carried out. For instance, if a read contains a single splice junction, a first seed is mapped to a donor splice site and the unmapped portion of the read is mapped again (in this case to an acceptor splice site). Subsequently, STAR builds

alignments of the entire read sequence by stitching together all the seeds that were aligned to the genome in the first step [17]. In our study, the parameters employed in STAR mapping were those reported by Zhang et al. [21] and the pig reference genome v. 10.2 (*Sscrofa* 10.2) was used as template.

Once reads were mapped, they were counted with the RSEM v. 1.3 [18] software by using default parameters with the option “-paired-end” and considering the porcine gene annotation file and the pig *Sscrofa* 10.2 genome sequence. RSEM generates a set of reference transcript sequences and subsequently a set of RNA-Seq reads are aligned to these reference transcripts [18]. Alignments generated with this procedure are used to infer transcript abundances by computing maximum likelihood abundance estimates with the Expectation-Maximization algorithm [18]. Credibility intervals at 95% are built with a Bayesian approach implemented in RSEM. Additional details can be found in Li et al. [18]. Read counts associated with each specific mRNA isoform were employed to carry out analysis of differential expression with DESeq2 [19]. DESeq2 assumes that read counts follow a negative binomial distribution, for each gene i and for each sample j , with a mean μ_{ij} and a dispersion value α_i . Means are proportional to the amounts of complementary deoxyribonucleic acid (cDNA) fragments corresponding to each gene scaled by a normalization factor. Gene-wise dispersion values are calculated with a maximum likelihood approach and subsequently they are shrunk towards a set of predicted dispersion values with an empirical Bayes approach. Subsequently, DESeq2 shrinks \log_2 fold-change (FC) estimates, with an empirical Bayes procedure [19], to reduce variance due to noisiness issues of genes that are poorly expressed. Finally, a Wald test is used to infer if shrunk \log_2 FC estimates (and their standard errors) are significantly different from zero. In the Wald test, the shrunken estimate of the \log_2 FC is divided by its standard error, generating a z -statistic that can be compared to a standard normal distribution [19].

Transcript annotation

To classify splicing events with the SUPPA [22] and Splicing Express [23] softwares, genome BAM files were generated with the STAR software [17], by using the same parameters described above. These BAM files were employed to assemble transcripts with Cufflinks [24], taking as a reference the *Sscrofa* 10.2 genome, and a master transcriptome was generated with Cuffmerge. The SUPPA software annotates AS events from a general input annotation file generated with Cuffmerge. The AS event types considered by SUPPA are: exon skipping, alternative 5' and 3' splice sites, and intron retention. For each event, SUPPA calculates the inclusion

parameter Ψ , which is defined as the ratio of the abundance of transcripts that include one form of the event over the abundance of the transcripts that contain either form of the event. On the other hand, the Splicing Express software uses a well-annotated set of reference sequences to detect different AS events from a transcriptome data (GTF file) input file i.e. exon skipping, intron retention and alternative 5' and 3' splicing borders. Splicing Express clusters expressed transcripts to identify their gene of origin and identifies AS events by using an algorithm based on the pairwise comparison. Besides, expressed sequences are represented as binary sequences (exons = 1, introns = 0) that are pairwise compared thus generating numerical patterns which reflect their splicing differences. Finally, a graphic representation of the expression level is created for each gene and for each identified AS event [23].

Transcript type annotation of porcine GM mRNA isoforms was retrieved from the BioMart database, available in the Ensembl database (<http://www.ensembl.org/biomart/martview/>). Gene Ontology (GO) Enrichment Analysis was performed by using the Panther database v. 12.0 (<http://www.pantherdb.org/>) with the data set of 87 genes simultaneously detected by both pipelines (P -value < 0.05) as producing mRNA isoforms DE in HIGH vs LOW pigs.

Validation of differentially expressed mRNA isoforms by RT-qPCR

Differential expression of mRNA isoforms was validated for the MAF BZIP transcription factor F (*MAFF*), stearoyl-CoA desaturase (*SCD*), retinoic acid receptor γ (*RXRG*) and integrin $\alpha 5$ (*ITGA5*) genes by reverse transcription-quantitative polymerase chain reaction (RT-qPCR). Primers spanning exon-exon boundaries, or alternatively binding at different exons (in order to avoid the amplification of residual contaminating genomic DNA), and complementary to exonic regions that define specific isoforms were designed with the Primer Express software (Applied Biosystems) (Additional file 2: Table S2). One μg of total RNA from 14 pigs (7 from each group - HIGH and LOW), selected at random from the global population of 52 pigs, was used as template for cDNA synthesis. The reverse transcription reaction was carried out with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA) in a final volume of 20 μl . Quantitative PCR reactions included 7.5 μl of SYBR Select Master Mix, 300 nM of each primer and 3.75 μl of a 1:25 dilution of the cDNA in a final volume reaction of 15 μl . Three genes e.g. β -actin (*ACTB*), TATA-Box binding protein (*TBP*) and hypoxanthine phosphoribosyltransferase 1 (*HPRT1*) were used as endogenous controls. The PCR thermal cycle involved one denaturing step at 95 °C for 10 min plus 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Reactions

were run in a QuantStudio 12 K Flex Real-Time PCR System (Applied Biosystems, Foster City, CA). A melting curve analysis i.e. 95 °C for 15 s, 60 °C for 15 s and a gradual increase in temperature, with a ramp rate of 1% up to 95 °C, followed by a final step of 95 °C for 15 s, was performed after the thermal cycling protocol to ensure the specificity of the amplification. We made sure that housekeeping and target genes had comparable amplification efficiencies (90–110%) by performing standard curve assays with serial 1:5 dilutions. Gene expression levels were quantified relative to the expression of endogenous controls by employing an optimized comparative Ct ($2^{-\Delta\Delta C_t}$ method) value approach [25] implemented in the Thermo Fisher Cloud (Thermo Fisher Scientific, Barcelona, Spain). Each sample was analysed in triplicate. All results were evaluated using RT-qPCR data analysis software (Thermo Fisher Cloud, Thermo Fisher Scientific, Barcelona, Spain). The sample displaying the lowest expression was used as calibrator. Differential expression was assessed with a Student's t-test.

Results

After quality control analysis, we used a final dataset of 52 GM muscle samples equally distributed between the HIGH ($N=26$) and LOW ($N=26$) groups. RNA-sequencing of these samples generated an average of 66 million paired-end reads per sample. The majority of reads (72.8%, CLC Bio; 89% STAR software) were successfully mapped to the pig *Sscrofa* 10.2 genome assembly. The mean mapping proportions obtained with CLC Bio were 91.4% and 8.6% for reads corresponding to exonic and intronic regions, respectively. When using STAR, 79.3% of the mapped reads were located in exons and 6.5% in introns. The remaining (14.2%) reads mapped to intergenic regions.

In the CLC Bio analysis, we found evidence of the existence of alternative transcripts in 2066 genes (11.7% of protein-coding genes expressed in the GM muscle of HIGH and LOW swine) which produced 5835 mRNA isoforms (2.8 transcripts per gene). In contrast, the STAR software detected 1430 genes (10.2% of expressed protein-coding genes) yielding 4391 different transcripts (3.0 transcripts per gene). Only 5.0% of alternative transcript variants were potentially subject to nonsense-mediated decay (Additional file 3: Table S3). Interestingly, 93% of the genes identified by STAR/RSEM/DESeq2 as displaying alternative transcripts were also detected with CLC Bio. Analysis of transcriptomic data with the SUPPA software [22] evidenced that exon skipping is the most prevalent AS event, while intron retention is the rarest one, i.e. they comprise 36.7% and 12.2% of all GM AS events, respectively (Additional file 4: Table S4). Similar results were obtained with the Splicing Express software [23]

(Additional file 4: Table S4), i.e. exon skipping was the most prevalent AS event (41.1%) and intron retention the least favoured one (12.7%).

We used two different pipelines (CLC Bio and STAR/RSEM/DESeq2) to detect DE mRNA isoforms in HIGH vs LOW pigs. Combination of such data sets made possible to identify 104 alternative transcripts and 87 genes that were simultaneously detected by both pipelines (P -value < 0.05) (Additional file 5: Table S5). A more stringent analysis (P -value < 0.01 and $\pm 0.6 \log_2 FC$) ascertained 10 DE transcripts (corresponding to 10 genes) concurrently discovered by both pipelines (Table 2). Five of these transcripts remained significant after correction for multiple testing (q -value < 0.05 and $\pm 0.6 \log_2 FC$; Table 2). In general, differential expression only affected one isoform and, more particularly, that showing a predominant pattern of expression (Tables 2, 3 and Additional file 6: Table S6). In order to validate the accuracy of our RNA-Seq approach, we measured the expression of four DE mRNA isoforms (*ITGA5*, *SCD*, *RXRG* and *MAFF*) by RT-qPCR analysis (Additional file 7: Figure S1). A significant differential expression was confirmed for two splicing variants e.g. *ITGA5* (4445 bp) and *SCD* (5585 bp) genes (P -value < 0.04). Besides, a strong statistical tendency was observed for the *RXRG* (544 bp) gene (P -value = 0.06). In contrast, the *MAFF* (2145 bp, P -value = 0.23) did not show a statistically significant differential expression, though RT-qPCR data reflected the same trends (FC and raw abundance estimates) detected by RNA-seq.

We carried out a GO analysis of the data set of 87 genes producing alternative transcripts (P -value < 0.05). We did not analyse the two other data sets (10 genes, P -value < 0.01 and $\pm 0.6 \log_2 FC$; 5 genes, q -value < 0.05 and $\pm 0.6 \log_2 FC$) because they are too small. The main molecular functions identified in the data set of 87 genes were *Binding* and *Catalytic activity* (Fig. 1a). These results are consistent with those of Lindholm et al. [26], who found that the main functions of mRNA encoding genes expressed in the human skeletal muscle are also related with binding and catalytic activity. The top GO terms of the cellular component GO category were *Membrane* and *Cell part* (Fig. 1b), while *Metabolism* and *Cellular process* were the most common biological processes amongst genes producing alternative transcripts (Fig. 1c). These results agree well with previous data obtained in humans, mouse and cow [23]. These functional processes are remarkably un-specific, thus probably reflecting the heterogeneous biological roles of genes expressing alternative transcripts.

Discussion

About 10.9% (average of CLC Bio and STAR results) of pig genes expressed in the GM muscle produced alternative transcripts, as opposed to 95% of genes

Table 2 Splicing variants that are differentially expressed (P -value < 0.01 and $\pm 0.6 \log_2$ Fold-Change) in the *gluteus medius* muscle of HIGH ($N = 26$) vs LOW pigs ($N = 26$)^a

Feature ID	Feature transcript ID	Transcript ID	Length (bp)	Relative expression mean (%)	CLC Bio		STAR/RSEM/DESeq2		q -value ^b	
					\log_2 (FC)	P -value	\log_2 (FC)	P -value		
ENSSSCG00000000293	ENSSSCT000000000314	ITGA5-201	4445	99.00	0.85	2.64E-05	2.56E-02	0.85	2.54E-06	1.21E-03
ENSSSCG000000001252	ENSSSCT000000001367	UBD-201	1281	75.81	-0.72	1.37E-03	2.29E-01	-0.83	7.73E-04	NA
ENSSSCG000000003079	ENSSSCT0000000003416	PVR-202	1167	48.41	0.77	7.76E-03	5.94E-01	0.66	9.27E-03	1.27E-01
ENSSSCG000000004578	ENSSSCT0000000005057	ANXA2-202	1455	99.36	0.70	3.32E-06	6.56E-03	0.69	4.56E-06	1.58E-03
ENSSSCG000000012277	ENSSSCT000000013426	TIMP1-001	931	97.85	2.57	4.61E-07	1.68E-03	0.79	2.01E-04	1.32E-02
ENSSSCG000000012653	ENSSSCT000000013834	ZDHH9-209	2967	17.79	0.77	3.29E-03	3.75E-01	0.77	2.56E-03	6.11E-02
ENSSSCG000000013579	ENSSSCT000000014831	CD209-001	1042	95.10	1.04	7.84E-05	4.18E-02	1.01	2.09E-05	NA
ENSSSCG000000006328	ENSSSCT0000000033501	RXRG-202	544	17.68	-0.70	4.09E-04	1.18E-01	-0.76	9.31E-05	8.98E-03
ENSSSCG000000009584	ENSSSCT0000000034286	SEMA4D-208	487	8.57	1.40	2.75E-06	6.53E-03	1.09	5.87E-05	6.59E-03
ENSSSCG000000024982	ENSSSCT0000000036552	LITAF-201	2190	88.44	1.01	1.02E-04	4.38E-02	0.79	1.21E-04	1.04E-02

^aDifferentially expressed mRNA isoforms that remained significant after correction for multiple testing (q -value < 0.05 and $\pm 0.6 \log_2$ Fold-Change) are shown in bold. A positive \log_2 FC means that the gene is upregulated in HIGH pigs

^bFor multiple testing correction, DESeq2 carries out a filtering step based on the average expression strength of each gene across all samples with the aim of discarding genes which are likely to lose significance after correcting for multiple testing. The purpose of this filtering step is to increase statistical power by reducing the list of candidate genes to be tested. The q -values of the genes which do not pass the filtering step are set to NA

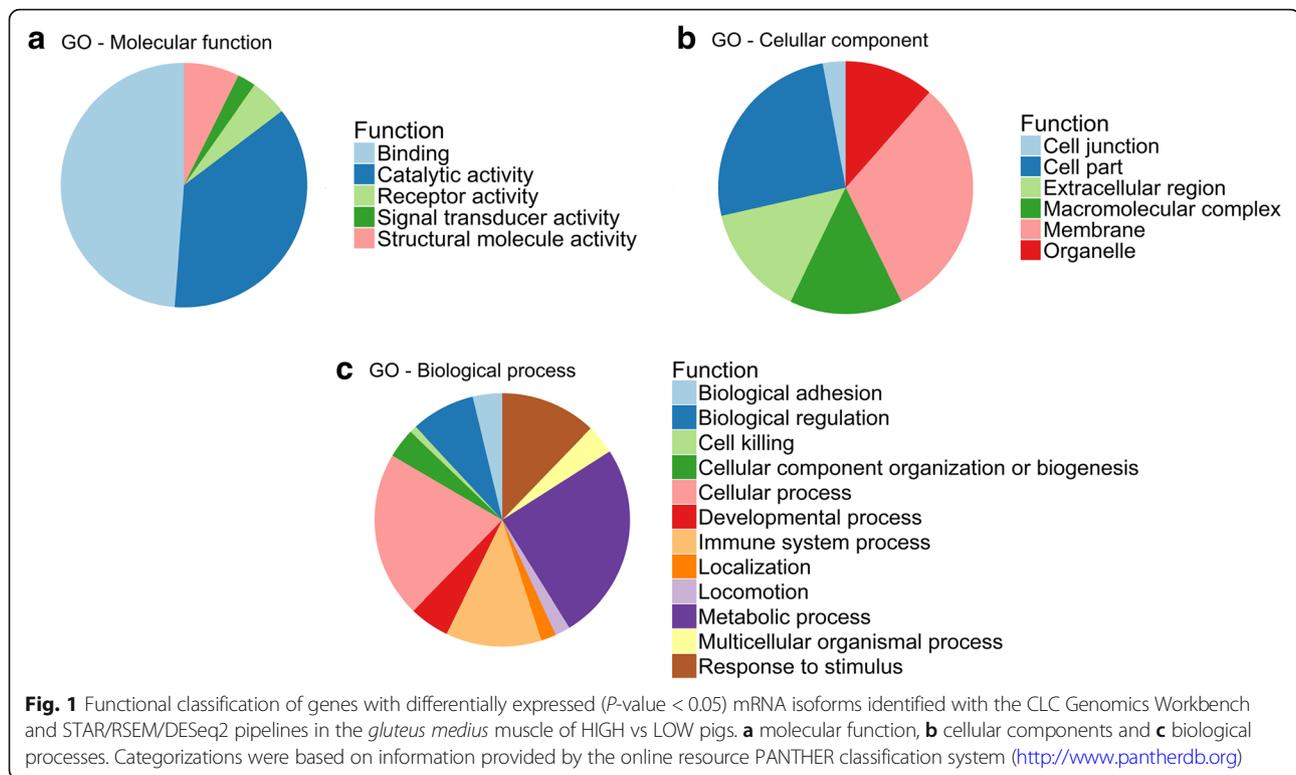
Table 3 Relative expression of the set of isoforms of four loci (*TIMP1*, *ITGA5*, *ANXA2* and *LITAF*) in HIGH (N = 26) vs LOW (N = 26) pigs^a

Feature ID	Feature transcript ID	Transcript ID	Length (bp)	Type ^b	Relative expression mean (%)	CLC Bio		STAR/RSEM/DESeq2			
						Log ₂ (FC)	P-value	Log ₂ (FC)	P-value	q-value ^c	
ENSSCG00000004578	ENSSSCT000000034155	ANXA2-201	1609	Protein coding	0.64	0.21	6.77E-01	1.00E+00	0.07	9.20E-01	NA
ENSSCG00000004578	ENSSSCT00000005057	ANXA2-202	1455	Protein coding	99.36	0.70	3.32E-06	7.23E-03	0.69	4.56E-06	1.58E-03
ENSSCG00000000293	ENSSSCT00000000314	ITGA5-201	4445	Protein coding	99.00	0.85	2.64E-05	2.82E-02	0.85	2.54E-06	1.21E-03
ENSSCG00000000293	ENSSSCT000000035821	ITGA5-203	1255	NMD	0.34	-0.03	9.18E-01	1.00E+00	-	-	-
ENSSCG00000000293	ENSSSCT000000034427	ITGA5-204	1013	NMD	0.71	-0.08	1.00E+00	1.00E+00	-	-	-
ENSSCG00000000293	ENSSSCT0000000033141	ITGA5-205	766	NMD	0.48	-0.29	1.00E+00	1.00E+00	-0.01	9.97E-01	NA
ENSSCG00000012277	ENSSSCT00000013426	TIMP1-001	931	Protein coding	97.86	2.57	4.61E-07	1.85E-03	0.79	2.01E-04	1.32E-02
ENSSCG00000012277	ENSSSCT000000033796	TIMP1-002	598	Protein coding	0.38	1.12	5.56E-01	1.00E+00	0.38	8.98E-01	NA
ENSSCG00000012277	ENSSSCT000000034602	TIMP1-003	641	Protein coding	1.57	1.19	8.47E-01	1.00E+00	-0.19	7.39E-01	NA
ENSSCG00000012277	ENSSSCT000000036308	TIMP1-004	173	Protein coding	0.19	3.05	4.21E-01	1.00E+00	0.18	9.52E-01	NA
ENSSCG00000024982	ENSSSCT000000036552	LITAF-201	2190	Protein coding	88.45	1.01	1.02E-04	4.82E-02	0.79	1.21E-04	1.04E-02
ENSSCG00000024982	ENSSSCT000000025103	LITAF-202	2370	Protein coding	11.55	-0.04	1.00E+00	1.00E+00	0.49	7.52E-01	NA

^aDifferentially expressed mRNA isoforms that remained significant after correction for multiple testing (q-value < 0.05 and ±0.6 log₂Fold-Change) are shown in bold. A positive log₂FC means that the gene is upregulated in HIGH pigs

^bNMD Nonsense-mediated mRNA decay

^cFor multiple testing correction, DESeq2 carries out a filtering step based on the average expression strength of each gene across all samples with the aim of discarding genes which are likely to lose significance after correcting for multiple testing. The purpose of this filtering step is to increase statistical power by reducing the list of candidate genes to be tested. The q-values of the genes which do not pass the filtering step are set to NA



detected in a broad array of human tissues [2]. Besides, the average number of mRNA isoforms per gene in the porcine skeletal muscle was 2.9 (average of CLC Bio and STAR results). The analysis of transcript diversity in the human skeletal muscle revealed a similar pattern, with an average of 2 isoforms per gene [26], a figure that is clearly below the average transcript diversity (5.4 isoforms/gene) found in other human tissues [27]. Such feature might be due to the fact that in the skeletal muscle there is a reduced number of transcripts (e.g. myofibrillar proteins) that encompass a disproportionate fraction of the total transcriptome [26]. Besides that, Taneri et al. [28] predicted a lower quantity of transcripts per gene in primary tissues e.g. skeletal muscle. In cattle, Chacko et al. [29] observed that 21% of genes are alternatively spliced, and similar percentages were observed by Kim et al. [30] in cows (26%) and dogs (14%).

The types of splicing events detected in the porcine muscle have been compiled in Additional file 4: Table S4. Exon skipping was the most frequent AS event predicted by SUPPA and Splicing Express softwares, followed by the use of alternative 5' splice (SUPPA) and 3' splice (Splicing Express) sites. Deep sequencing of 15 human tissues and cell lines was consistent with these findings, thus demonstrating that exon skipping was the most frequent AS event, followed by the use of alternative 3' splice and 5' splice sites [31]. In accordance with results obtained in

humans [31], the less frequent event was intron retention (Additional file 4: Table S4). Besides, only 5.0% of produced transcripts were predicted to undergo nonsense-mediated decay. In humans, nonsense-mediated decay and nuclear sequestration and turnover of intron-retention transcripts have been involved in the downregulation of genes in tissues where they do not have a relevant physiological role [32]. Though SUPPA and Splicing Express yielded consistent results about the relative importance of distinct AS event categories in the porcine skeletal muscle, the sets of genes identified by these two softwares as yielding alternative transcripts were quite different. The proportions of genes overlapping the SUPPA and Splicing Express data sets classified according to the type of AS event were: exon skipping = 27%, alternative 5' = 24%, 3' splice sites = 20% and intron retention = 14%. Though we do not have a straightforward explanation for these discrepancies, we hypothesize that they might be due to the existence of relevant differences in the assumptions and algorithms on which these two softwares are based.

Estimating isoform mRNA abundance is a challenging task and results may vary depending on the bioinformatics approach employed in differential expression analysis [33]. One of the main factors influencing the outcome of differential mRNA isoform expression is the quality and completeness of the transcript assembly [33]. The mRNA isoform annotation of the pig genome

is still incomplete and obviously this might affect the results of our analysis, but we have not attempted to reconstruct transcripts because bioinformatic and statistical approaches to do so are not very robust and they may lead to inaccurate transcript quantitations [34].

In order to obtain results as much precise as possible, we have used two different pipelines to identify DE mRNA isoforms and we have considered as genuine differential expression events those identified by both approaches. This combined analysis highlighted the existence of five genes with DE mRNA isoforms that remained significant after correction for multiple testing (q -value < 0.05 , $\pm 0.6 \log_2 FC$). It is worth to highlight that the DE isoform (487 bp) of the pig semaphorin 4D (*SEMA4D*) gene is annotated, in the Ensembl database (*Sscrofa* 10.2 assembly; <https://www.ensembl.org>), as truncated in its 3'end. In the human *SEMA4D* gene, there are 13 protein-encoding mRNA isoforms and five of them are also truncated in their 3'ends (*GRCh38.p10* assembly; <https://www.ensembl.org>). The existence of truncated transcripts is due to the inability of conventional RNA-seq experiments to define the ends of genes with high precision [35]. Moreover, automated gene prediction is a difficult task and, in consequence, first-pass annotations can be quite inaccurate [35]. Obviously, the analysis of the differential expression of mRNA isoforms strongly depends on the accuracy of transcript annotation, so the results presented in the current work need to be interpreted with this caveat in mind.

By examining Tables 2 and 3, we have noticed that the differential expression of mRNA isoforms might have different functional consequences depending on the gene under consideration. For instance, in the case of the *ITGA5* and TIMP metalloproteinase inhibitor 1 (*TIMP1*) genes the DE mRNA isoform encodes a protein that is longer than the proteins encoded by the remaining mRNA isoforms recorded in the Ensembl database (*Sscrofa* 10.2 assembly; <https://www.ensembl.org>). With regard to *ITGA5*, the DE isoform (4445 bp) encodes a full length protein of 1057 amino acids (aa) and it has a predominant pattern of expression (99%), while the remaining porcine *ITGA5* isoforms reported in the Ensembl database correspond to processed transcripts or transcripts subject to nonsense mediated decay. Similarly, in humans there is one major *ITGA5* isoform (4444 bp), another one that might encode a protein but it is truncated in its 5'end, and eleven isoforms that correspond to processed transcripts, retained introns and transcripts subject to nonsense mediated decay (<https://www.ensembl.org>). Concerning the *TIMP1* gene, the DE isoform (931 bp) encodes a full length protein of 207 aa that is longer than the proteins encoded by other isoforms (*Sscrofa* 10.2 assembly; <https://www.ensembl.org>): 195 aa (but incomplete 5'end), 123 aa and 38 aa (but

incomplete 3'end). If we compare the 207 aa (931 bp transcript) and the 123 aa (598 bp transcript) *TIMP1* isoforms, the latter lacks a central part of the protein (from aa site 68 to 151), a feature that involves the loss of four of the six disulfide bridges which stabilize the fold of the molecule and of two aa residues (sites 68 and 69) which bind to the catalytic zinc [36, 37]. These observations imply that the two 207 aa and 123 aa porcine protein isoforms are expected to be very different at the functional level.

A different case is represented by the annexin A2 (*ANXA2*) and lipopolysaccharide-induced TNF-alpha factor (*LITAF*) genes in which the DE mRNA isoform (*ANXA2*: 1455 bp, *LITAF*: 2190 bp) is shorter than the longest annotated transcript (*ANXA2*: 1609 bp, *LITAF*: 2370 bp, Table 3) but both encode proteins of identical length (*ANXA2*: 339 aa, *LITAF*: 161 aa, *Sscrofa* 10.2 assembly; <https://www.ensembl.org>). This situation is comparable to what has been reported in humans for the *ANXA2* (1676 bp, 1444 bp and 1435 bp isoforms encoding a protein of 339 aa) and *LITAF* (six different isoforms e.g. 2632 bp, 2467 bp, 2356 bp, 1118 bp, 717 bp and 603 bp mRNAs encoding a protein of 161 aa) genes (*GRCh38.p10* assembly; <http://www.ensembl.org>). Though proteins with an identical length and sequence composition should be functionally equivalent, differences in transcript length might affect mRNA translatability (e.g. presence of short upstream open reading frames in the 5'UTR), stability (e.g. formation of stable stem-loops, presence of microRNA binding sites and of AU-rich elements) and cell localization [38].

The upregulation of certain mRNA isoforms of the *ITGA5*, *TIMP1*, *ANXA2* and *LITAF* genes in HIGH pigs is relevant because these four genes have been implicated in human obesity and diabetes. For instance, high glucose concentrations induce the overexpression of the fibronectin receptor, an heterodimer whose α -chain is encoded by the *ITGA5* gene [39]. Moreover, *TIMP1* expression is increased in the serum and adipose tissue of obese mouse models [40]. There is also evidence that the knockout of the *ANXA2* gene in mice involves an hypotrophy of the white adipose tissue due to reduced fatty acid uptake [41], and *LITAF* mRNA is overexpressed in overweight and obese humans [42]. In summary, the upregulation of these four genes in HIGH swine is consistent with the increased fatness and live weight of these pigs and suggests that the differential expression of specific mRNA isoforms might contribute to the phenotypic differences observed in HIGH vs LOW pigs.

Finally, we would like to discuss a third case in which DE mRNA isoforms encode proteins that are shorter than the canonical full-length protein. We have observed that a 3155 bp transcript corresponding to the porcine

ubiquitin specific peptidase 2 (*USP2*) gene and encoding a 396 aa protein is upregulated in HIGH pigs (Additional file 5: Table S5). In the Ensembl database (<http://www.ensembl.org>, *Sscrofa* 10.2), a second mRNA isoform that encodes a 606 aa protein has been annotated. In humans, two USP2 protein isoforms of 605 aa (USP2–69) and 396 aa (USP2–45) have been reported and that there are evidences that both are able to prevent the degradation of the low density lipoprotein (LDL) receptor. The upregulation of the 396 aa *USP2* isoform in HIGH swine might constitute a mechanism to cope with the elevated serum LDL concentrations observed in this group of pigs (Table 1). It is also worth to highlight that the USP2–69 and USP2–45 isoforms might not be functionally equivalent. In *Xenopus*, for instance, USP2–45 can deubiquitylate epithelial Na⁺ channels in oocytes, while USP2–69 cannot perform such function due to differences in their N-terminal domains. In humans, functional differences have been also observed with regard to the implication of *USP2* isoforms in cell cycle progression and antiviral response [43], but unfortunately no such data are currently available for pigs.

Conclusions

We have demonstrated that around 10.9% of genes expressed in the porcine skeletal muscle produce alternative transcripts, thus generating an average of 2.9 different mRNA isoforms per gene. Exon skipping is the most frequent splicing event, followed by the use of alternative 5' splice sites (SUPPA) and 3' splice sites (Splicing Express). By analysing the differential expression of mRNA isoforms in HIGH vs LOW pigs, we have demonstrated that in the GM muscle of HIGH pigs, which display an increased fatness, specific *ITGA5*, *ANXA2*, *LITAF* and *TIMP1* mRNA isoforms are upregulated. This finding is biologically meaningful because these four genes have been implicated in human obesity and metabolism [39–42]. A deeper functional characterization of these mRNA isoforms, through initiatives such as the Functional Annotation of Farm Animal Genomes project [44], will be essential to infer the consequences of their differential expression on porcine growth and fatness.

Additional files

Additional file 1: Table S1. Distribution of the 56 animals sequenced by RNA-seq in the 5 half-sib families reported by Gallardo et al. [13]. (XLSX 9 kb)

Additional file 2: Table S2. Primers employed in the validation of four differentially expressed isoforms by RT-qPCR. (XLSX 10 kb)

Additional file 3: Table S3. Alternatively spliced mRNA isoforms identified in the porcine *gluteus medius* muscle of Duroc pigs by CLC Bio and/or STAR/RSEM/DESeq2. (XLSX 569 kb)

Additional file 4: Table S4. Classification of alternative splicing (AS) events detected in the porcine *gluteus medius* muscle with the SUPPA and Splicing Express softwares. (XLSX 11 kb)

Additional file 5: Table S5. Differentially expressed (*P*-value < 0.05) mRNA isoforms (HIGH vs LOW pigs) found with CLC Bio and STAR/RSEM/DESeq2 softwares. (XLSX 225 kb)

Additional file 6: Table S6. Relative transcript levels of a set of isoforms corresponding to five genes expressed in the *gluteus medius* muscle of HIGH and LOW pigs identified with the CLC Bio and STAR/RSEM/DESeq2 pipelines. (XLSX 160 kb)

Additional file 7: Figure S1. Validation by RT-qPCR of the differential expression of mRNA isoforms corresponding to the *RXRG*, *SCD*, *MAFF* and *ITGA5* genes in HIGH vs LOW pigs. (PPT 132 kb)

Abbreviations

aa: Amino acids; *ACTB*: β -actin; *ANXA2*: Annexin A2; cDNA: Complementary deoxyribonucleic acid; DE: Differentially expressed; FC: Fold-change; GBP5: Guanylate binding protein 5; GM: *Gluteus medius*; GO: Gene ontology; *HPRT1*: Hypoxanthine phosphoribosyltransferase 1; *ITGA5*: Integrin $\alpha 5$; LDL: Low density lipoprotein; LGM: Large Gap Mapper; *LITAF*: Lipopolysaccharide-induced TNF- α factor; *MAFF*: MAF BZIP transcription factor F; MUFA: Monounsaturated fatty acids; PUFA: Polyunsaturated fatty acids; RNA-seq: RNA-sequencing; RSEM: RNA-Seq by Expectation Maximization software; RT-qPCR: Reverse transcription-quantitative polymerase chain reaction; *RXRG*: Retinoic acid receptor γ ; *SCD*: Stearoyl-CoA desaturase; *SEMA4D*: Semaphorin 4D; SFA: Saturated fatty acids; STAR: Spliced Transcripts Alignment to a Reference Software; *TBP*: TATA-Box binding protein; *TIMP1*: TIMP metalloproteinase inhibitor 1; *USP2*: Ubiquitin specific peptidase 2

Acknowledgements

The authors are indebted to Selecció Batallé S.A. for providing the animal material. We gratefully acknowledge to J. Reixach (Selecció Batallé), J. Soler (IRTA), C. Millan (IRTA), A. Quintana (IRTA) and A. Rossell (IRTA) for their collaboration in the experimental protocols and pig management.

Funding

Part of the research presented in this publication was funded by grants AGL2013–48742-C2–1-R and AGL2013–48742-C2–2-R awarded by the Spanish Ministry of Economy and Competitiveness and grant 2014 SGR 1528 from the Agency for Management of University and Research Grants of the Generalitat de Catalunya. We also acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the *Center of Excellence Severo Ochoa 2016–2019* (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG). Tainã Figueiredo Cardoso was funded with a fellowship from the CAPES Foundation–Coordination of Improvement of Higher Education, Ministry of Education of the Federal Government of Brazil. Rayner Gonzalez-Prendes was funded with a FPU Ph.D. grant from the Spanish Ministry of Education (FPU12/00860). Thanks also to the CERCA Programme of the Generalitat de Catalunya. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Availability of data and materials

Data have been deposited in the Figshare database (<https://doi.org/10.6084/m9.figshare.5831235.v1>).

Authors' contributions

MA, ACan and RQ designed the experiment; RQ was responsible for the experimental protocols and generation of animal material; all authors contributed to the obtaining of biological samples; TFC and RGP performed RNA extractions; TFC analysed the RNA-sequencing data; ACAs designed the RT-qPCR experiments; TFC carried out the RT-qPCR validation experiments; TFC and ACAs analysed the RT-qPCR data; MA, ACan and TFC wrote the paper. All authors read and approved the manuscript.

Ethics approval and consent to participate

Animal care, management procedures and blood sampling were performed following national guidelines for the Good Experimental Practices and they were approved by the Ethical Committee of the Institut de Recerca i Tecnologia Agroalimentàries (IRTA).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. ²CAPES Foundation, Ministry of Education of Brazil, Brasília D.F. 70.040-020, Brazil. ³Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimón, 08140 Caldes de Montbui, Spain. ⁴Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. ⁵Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada.

Received: 23 December 2016 Accepted: 31 January 2018

Published online: 14 February 2018

References

- Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 2017;46:582–92.
- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, et al. Deciphering the splicing code. *Nature.* 2010;465:53–9.
- Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, et al. RNA sequencing of cancer reveals novel splicing alterations. *Sci Rep.* 2013;3:93–7.
- Yao Y, Shang J, Song W, Deng Q, Liu H, Zhou Y. Global profiling of the gene expression and alternative splicing events during hypoxia-regulated chondrogenic differentiation in human cartilage endplate-derived stem cells. *Genomics.* 2016;107:170–7.
- Zhang S, Cai H, Yang Q, Shi T, Pan C, Lei C, et al. Identification of novel alternative splicing transcript and expression analysis of bovine *TMEM95* gene. *Gene.* 2016;575:531–6.
- Xie Y, Yang S, Cui X, Jiang L, Zhang S, Zhang Q, et al. Identification and expression pattern of two novel alternative splicing variants of *EEF1D* gene of dairy cattle. *Gene.* 2014;534:189–96.
- Ma J, Yang J, Zhou L, Ren J, Liu X, Zhang H, et al. A splice mutation in the *PHKG1* gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLoS Genet.* 2014;10:e1004710.
- Koltes JE, Fritz-Waters E, Easley CJ, Choi I, Bao H, Kommadath A, et al. Identification of a putative quantitative trait nucleotide in guanylate binding protein 5 for host response to PRRS virus infection. *BMC Genomics.* 2015;16:412.
- Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, et al. Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. *PLoS One.* 2014;9:e99720.
- Cánovas A, Quintanilla R, Amills M, Pena RN. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics.* 2010;11:372.
- Ayuso M, Óvilo C, Rodríguez-Bertos A, Rey AI, Daza A, Fenández A, et al. Dietary vitamin A restriction affects adipocyte differentiation and fatty acid composition of intramuscular fat in Iberian pigs. *Meat Sci.* 2015;108:9–16.
- Cánovas A, Pena RN, Gallardo D, Ramírez O, Amills M, Quintanilla R, et al. Segregation of regulatory polymorphisms with effects on the *gluteus medius* transcriptome in a purebred pig population. *PLoS One.* 2012;7:e35583.
- Gallardo D, Pena RN, Amills M, Varona L, Ramírez O, Reixach J, et al. Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations in pigs. *Physiol Genomics.* 2008;35:199–209.
- Mach N, Devant M, Díaz I, Font-Furnols M, Oliver MA, García JA, et al. Increasing the amount of n-3 fatty acid in meat from young Holstein bulls through nutrition. *J Anim Sci.* 2006;84:3039–48.
- Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem.* 1972;18:499–502.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
- Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics.* 2008;9:321–32.
- Zhang C, Zhang B, Lin LL, Zhao S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC Genomics.* 2017;18:583.
- Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyras E. Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA.* 2015;21:1521–31.
- Kroll JE, Kim J, Ohno-Machado L, de Souza SJ. Splicing Express: a software suite for alternative splicing analysis using next-generation sequencing data. *PeerJ.* 2015;3:e1419.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28:511–5.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001;25:402–8.
- Lindholm ME, Huss M, Solnestam BW, Kjellqvist S, Lundeberg J, Sundberg CJ. The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J.* 2014;28:4571–81.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007;447:799–816.
- Taner B, Snyder B, Novorodovsky A, Gaasterland T. Alternative splicing of mouse transcription factors affects their DNA-binding domain architecture and is tissue specific. *Genome Biol.* 2004;5:R75.
- Chacko E, Ranganathan S. Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. *BMC Genomics.* 2009;10:S3–S11.
- Kim N, Alekseyenko AV, Roy M, Lee C. The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res.* 2007;35:D93–8.
- Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008;456:470–6.
- Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gontopoulos-Pournatzis T, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 2014;24:1774–86.
- Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. *Nucleic Acids Res.* 2016;44:e127.
- Shenker S, Miura P, Sanfilippo P, Lai EC. IsoSCM: improved and alternative 3' UTR annotation using multiple change-point inference. *RNA.* 2015;21:14–27.
- Schurch NJ, Cole C, Sherstnev A, Song J, Duc C, Storey KG, et al. Improved annotation of 3' untranslated regions and complex loci by combination of strand-specific direct RNA sequencing, RNA-Seq and ESTs. *PLoS One.* 2014;9:e94270.
- Batra J, Robinson J, Soares AS, Fields AP, Radisky DC, Radisky ES. Matrix metalloproteinase-10 (*MMP-10*) interaction with tissue inhibitors of metalloproteinases *TIMP-1* and *TIMP-2*: binding studies and crystal structure. *J Biol Chem.* 2012;287:15935–46.
- Gomis-Rüth F-X, Maskos K, Betz M, Bergner A, Huber R, et al. Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by *TIMP-1*. *Nature.* 1997;389:77–81.

38. Mockenhaupt S, Makeyev EV. Non-coding functions of alternative pre-mRNA splicing in development. *Semin Cell Dev Biol.* 2015;47–48:32–9.
39. Roth T, Podestá F, Stepp MA, Boeri D, Lorenzi M. Integrin overexpression induced by high glucose and by human diabetes: potential pathway to cell dysfunction in diabetic microangiopathy. *Proc Natl Acad Sci USA.* 1993;90:9640–4.
40. Meissburger B, Stachorski L, Röder E, Rudofsky G, Wolfrum C. Tissue inhibitor of matrix metalloproteinase 1 (*TIMP1*) controls adipogenesis in obesity in mice and in humans. *Diabetologia.* 2011;54:1468–79.
41. Salameh A, Daquinag AC, Staquicini DI, An Z, Hajjar KA, Pasqualini R, et al. Prohibitin/annexin 2 interaction regulates fatty acid transport in adipose tissue. *JCI Insight.* 2016;1:e86351.
42. Ji ZZ, Dai Z, Xu YC. A new tumor necrosis factor (TNF)- α regulator, lipopolysaccharides-induced TNF- α factor, is associated with obesity and insulin resistance. *Chin Med J.* 2011;124:177–82.
43. Zhu H-Q, Gao F-H. The molecular mechanisms of regulation on USP2's alternative splicing and the significance of its products. *Int J Biol Sci.* 2017; 13:1489–96.
44. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the functional annotation of animal genomes project. *Genome Biol.* 2015;16:57.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

