# An exon-intron split framework to prioritize transcriptional and post-transcriptional regulatory signals and its application to study energy homeostasis in pigs

Emilio Mármol-Sánchez[1*], Susanna Cirera[2], Laura Zingaretti[1], Mette Juul Jacobsen[2], Yuliaxis Ramayo-Caldas[3], Claus B Jørgensen[2], Merete Fredholm[2], Tainã Figueiredo Cardoso[1†], Raquel Quintanilla[3] and Marcel Amills[1,4]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.

[2]Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, 1871 Frederiksberg C, Denmark.

[3]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, 08140 Caldes de Montbui, Spain.

[4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

*Emilio Mármol-Sánchez current address: Stockholm University, The Wenner-Gren Institute, Department of Molecular Biosciences, SciLifelab.

[†]Tainã Figueiredo Cardoso current address: Embrapa Pecuária Sudeste, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), 13560-970, São Carlos, SP, Brazil.

**Corresponding author:** Emilio Mármol-Sánchez. Stockholm University, The Wenner-Gren Institute, Department of Molecular Biosciences, SciLifelab. Email: emilio.marmol.sanchez@gmail.com

# Abstract

Bulk sequencing of RNA transcripts has typically been used to quantify gene expression levels in different experimental systems. However, linking differentially expressed (DE) mRNA transcripts to gene expression regulators, such as miRNAs or transcription factors (TFs), remains challenging, as in silico or experimental interactions are commonly identified post hoc after selecting differentially expressed genes of interest, thus biasing the interpretation of underlying gene regulatory mechanisms.

In this study, we performed an exon-intron split analysis (EISA) to muscle and fat RNA-seq data from two Duroc pig populations subjected to fasting-feeding conditions and with divergent fatness profiles, respectively. We compared the number of reads from exonic and intronic regions for all expressed protein-coding genes and divided their expression profiles into transcriptional and post-transcriptional components, considering intronic and exonic fractions as estimates of the abundance of pre-mRNA and mature mRNA transcripts, respectively. In this way, we obtained a prioritized list of genes showing significant transcriptional and post-transcriptional regulatory signals.

After running EISA analyses, protein-coding mRNA genes with downregulated exonic fractions and high post-transcriptional signals were significantly enriched for binding sites of upregulated DE miRNAs. Moreover, these genes showed an increased expression covariation for the exonic fraction compared to that of the intronic fraction. On the contrary, they did not show enrichment for binding sites of highly expressed and/or downregulated DE miRNAs. Among the set of loci displaying miRNA-driven post-transcriptional regulatory signals, we observed genes related to glucose homeostasis (*PDK4*, *NR4A3*, *CHRNA1* and *DKK2*), cell differentiation (*MYO9A*, *KLF5* and *BACH2*) or adipocytes metabolism (*LEP*, *SERPINE2*, *RNF157*, *OSBPL10* and *PRSS23*).

Besides, genes showing upregulated intronic fractions with a lack of exonic fractions were significantly enriched for TF-enhancer activity while depleted for miRNA targets, thus suggesting a transient transcription activation regulating skeletal muscle development. Our results highlight an efficient framework to classify mRNA genes showing transcriptional and post-transcriptional signals linked to transient transcription and miRNA-driven downregulation by using exonic and intronic fractions of RNA-seq datasets from muscle and adipose tissues in pigs.

**Keywords:**

Exon-intron split analysis, microRNA, Transcription factor, pigs, energy homeostasis.

# Background

RNA dynamics in the cell metabolism are subjected to complex yet poorly characterized regulatory mechanisms that contribute to shaping fine-tuned biological responses to different stimuli [1]. Cellular metabolic changes are hence a direct manifestation of intricate interactions between expressed transcripts and other regulatory elements that modify their abundance, localization, fate and degradation rate. Transcription factors (TFs) and microRNAs (miRNAs) are two of the major regulatory factors that alter RNA function at the transcriptional and post-transcriptional levels, respectively [2, 3] .Both trigger changes in the abundance of targeted transcripts and proteins, which can ultimately be modelled as positive or negative covariation patterns that might help unravel direct or indirect molecular interplays regulating biological pathways.

TFs are a particular class of proteins that bind to specific DNA motifs, facilitating the recruitment or blockage of other certain effectors, hence promoting or limiting the expression of the targeted genes under their regulatory influence [4]. Moreover, miRNAs are a class of small non-coding RNAs (~22 nucleotide long) mainly devoted to the post-transcriptional control of gene expression through translational inhibition and/or targeted mRNAs destabilization through poly(A) shortening and subsequent degradation [5].

In order to disentangle regulatory functions driven by TF-binding or miRNAs, researchers usually focus on genes showing changes in mRNA abundance or protein levels that such regulatory effectors might putatively target. This approach, however, introduces *post hoc* selection of deregulated genes and *ad hoc* search of predicted interactions by TF-motif enrichment in promoter regions or seed matches in 3'-UTRs for miRNAs. Such a bias in which genes are to be analyzed could obscure other mild yet relevant regulatory interactions that might not arise from differential expression analyses alone, and that

could be responsible of true occurring regulatory interactions within the cell metabolism [3, 6, 7].

Several studies have addressed the effects of transcriptionally and post-transcriptionally perturbed genes in expression datasets [8–10] or how transcriptional dynamics can reflect the cellular transition between homeostasis and stress-induced responses or differentiation stages [11–14]. In an attempt to capture both regulatory components based on expression data, Gaidatzis *et al*. [8] described the use of intronic mapping sequences as direct indicators of primary mRNA oscillations related to transcriptional activation. This study was based on early reports describing intronic expression as a proxy of nascent transcription and co-transcriptional splicing events [15–17]. Besides, the post-transcriptional effect was defined as a function of the expressed exonic and intronic fraction for each gene. Modelling the exonic and intronic contribution to gene expression changes between conditions, these and other authors could discern transcriptional and post-transcriptional responses within co-expressed mRNAs and their corresponding putative regulators [8–10, 18]. Nevertheless, despite these promising results, the use of such methodologies to identify relevant regulatory effects in high throughput experiments is still limited.

In the present study, we aimed to elaborate on such approaches in order to characterize the transcriptional and post-transcriptional responses in skeletal muscle in response to nutrient boost after food intake. Furthermore, we also explored post-transcriptional regulatory signals in adipose tissue using an independent pig population with divergent fatness profiles and qPCR analyses for cross-validation. Using a set of expression data from RNA-seq and miRNA-seq experiments in the two aforementioned porcine populations, we predicted in silico miRNA-mRNA and TF-DNA interactions from two different temporary feeding stages: (i) fasting vs. 7h after nutrient supply in muscle and

(ii) from pigs with high and low body mass index (BMI) profiles in adipocytes. We then evaluated the perturbed expression status of relevant genes at the transcriptional and/or post-transcriptional level in order to infer hidden regulatory effects contributing to the observed differences in gene expression between groups.

# Materials and methods

## Experimental design and tissue collection

Expression data and experimental conditions employed to infer gene covariation networks and regulatory connections were previously described in [19–21].

In brief, two independent experimental designs comprising expression profiles from mRNAs and miRNAs in Duroc pigs were used:

(i) The transcriptomic profile of *gluteus medius* (GM) skeletal muscle samples from a total of 23 Duroc gilts were measured employing RNA-seq and small RNA-seq sequencing using the same biological material as reported in [19, 21, 22]. All gilts were fed *ad libitum* during their productive lives [19, 22] until reaching ~150 days of age and they were subsequently slaughtered at the IRTA Experimental Slaughterhouse in Monells (Girona, Spain) following Spanish welfare regulations. Prior to slaughtering, all animals were fasted during 12 h. Gilts were then divided in two fasting/feeding regimes, *i. e.* 11 gilts (*AL-T0*) were slaughtered in fasting conditions upon arrival at the abattoir, whereas the rest of the animals were slaughtered immediately after 7 h (*AL-T2*, N =12) with access to *ad libitum* feed intake. After slaughtering, *gluteus medius* skeletal muscle samples were collected, conserved in RNAlater solution (Thermo Fisher Scientific, Barcelona, Spain) and subsequently snap-frozen in liquid nitrogen.

(ii) A total of 10 Duroc-Göttingen minipig inter-cross F2 animals with divergent fatness profiles according to BMI metric (5 *lean* and 5 *obese*) were selected from the UNIK resource population comprising a total of 502 F2 pigs [23, 24], as described in Jacobsen et al. 2019 [20]. Pigs were slaughtered when they reach 8-13 months of age according to the Danish "Animal Maintenance Act" (Act 432 dated 09/06/2004). Tissue samples from retroperitoneal tissue of each animal were collected and mature adipocytes were isolated following the protocol of Decaunes et al. 2011 [25] with modifications as reported in [20]. Once extracted, adipocyte pellets were snap-frozen at -80°C until further experimental procedures.

**RNA extraction and sequencing**

RNA extraction, RNA-Seq and small RNA-Seq sequencing protocols, quality-check preprocessing and mapping were performed as previously described [19–21]. In brief, total RNA was isolated from GM tissue using the RiboPure kit (Ambion, Austin, TX) and following the manufacturer's recommendations, while for RNA from adipocytes, the protocol from Cirera et al. 2013 [26] specifically adapted to adipose tissue was implemented. Sequencing libraries were prepared with dedicated TruSeq stranded total RNA kits (Illumina Inc. CA) [20, 21] and paired-end sequenced in a HiSeq 2000 equipment (Illumina Inc. CA). Small RNA-specific libraries were prepared from total RNA isolates following the TruSeq Small RNA Sample Preparation Kit (Illumina Inc., CA) according to the protocols of the manufacturer and subjected to single-end (1 x 50 bp) sequencing in a HiSeq 2500 equipment (Illumina Inc., CA).

**Quality check, filtering and mapping of sequences**

Sequenced reads from RNA-Seq and small RNA-Seq data sets belonging to both fasting and fed Duroc gilts, as well as to lean and obese Duroc-Göttingen minipigs, were trimmed to remove any remaining sequencing adapters and quality-checked with the Cutadapt software [27]. Subsequently, reads were mapped against the Sscrofa11.1 porcine reference assembly with the HISAT2 aligner [28] and default parameters for RNA-Seq reads, and with the Bowtie Alignment v.1.2.1.1 software [29] using small sequence reads specifications (*bowtie -n 0 -l 25 -m 20 -k 1 --best --strata*) for small RNA-Seq reads, respectively.

**Exon/Intron quantification**

To account for the separate determination of the exonic and intronic fractions for each annotated mRNA gene, the gtf formatted Sscrofa.11.1 v. 103 gene annotation file was retrieved from Ensembl repositories (http://ftp.ensembl.org/pub/release-103/gtf/sus_scrofa/) and processed to generate exonic and intronic ranges spanning all gene annotations available. Overlapping intronic/exonic regions, as well as singleton positions were removed from intronic ranges implementing custom scripts and functions available in GenomicFeatures and IRanges R modules [30]. To avoid the counting of exonic reads mapping close to exon/intron junctions that could bias the quantification of intronic regions, we removed 10 basepairs (bp) from both sides of each intronic range.

Once the corresponding exonic and intronic ranges for Sscrofa11.1 v. 103 were retrieved, the featureCounts function within the Rsubread package [31] was used to quantify gene expression profiles based on exon/intron expression patterns for each gene using paired-end non-overlapping specifications.

Moreover, miRNA expression profiles were extracted from small RNA-Seq sequencing data using the Sscrofa11.1 v103 mature miRNA annotation with featureCounts software tool [32] in single-end mode and with default parameters.

**Exon/intron split analysis (EISA) for assessing transcriptional and post-transcriptional effects on gene expression**

We applied EISA analyses differentiate gene expression regulation based on transcriptional and post-transcriptional effects [8–10]. To this end, the exonic/intronic abundance of each annotated mRNA gene was estimated by separately assigning mapping reads from RNA-Seq data sets using Sscrofa11.1 v103 exon/intron custom annotation ranges previously generated. Only genes showing average expression values above 1 count-per-million (CPM) in at least 50% of animals were retained for further analyses.

Counts were processed following the methods described in Gaidatzis *et al*. [8], where normalization was performed independently for exon and intron counts by multiplying each $i^{th}$ gene expression in each $j^{th}$ sample by the corresponding mean gene expression and dividing by the total number of quantified counts per sample.

Normalized and expression filtered gene abundances for exonic and intronic ranges were subsequently transformed in $\log_2$ scale, adding a pseudo-count of 1. Exon/intron $\log_2$ transformed abundances for each gene were averaged within each considered treatment groups (*AL-T0* and *AL-T2* for GM tissues and *lean* and *obese* for adipocyte isolates).

Only genes with successful exonic and intronic quantified read counts were further considered in our analyses, hence removing mono-exonic genes and poliexonic genes where exonic and/or intronic expression was undetected. The transcriptional component ($T_c$) contribution to observed differences in each $i^{th}$ gene expression levels between contrast groups was expressed as the increment of intron counts in fed (*AL-T2*) and *obese*

animals with respect to fasting (*AL-T0*) and *lean* animals, respectively, denoted as $\Delta Int = Int_{2i} - Int_{1i}$. The increment of exonic counts $\Delta Ex$ was defined accordingly and the post-transcriptional component ($PT_c$) effect was expressed as $\Delta Ex - \Delta Int = (Ex_{2i} - Ex_{1i}) - (Int_{2i} - Int_{1i})$. Finally, both components were z-scored to represent comparable ranges between $\Delta Ex$ and $\Delta Int$ estimates. Transcriptional and post-transcriptional expected effects according to $\Delta Ex$ and $\Delta Int$ distribution are depicted in **Fig. 1a-c**. The classification and interpretation of the combinatorial possibilities among $\Delta Ex$, $\Delta Int$ ($T_c$) and $\Delta Ex - \Delta Int$ ($PT_c$) values are shown in **Fig. 1d**. All implemented steps and instructions for running EISA analyses have been summarized in a ready-to-use modular pipeline publicly available at https://github.com/emarmolsanchez/EISACompR.

**Transcriptional and post-transcriptional signal prioritization**

In order to obtain a prioritized list of genes showing meaningful signals of transcriptional and post-transcriptional regulatory activity, the top 1% and 5% genes with the highest negative $PT_c$ scores were retrieved for each of the two experimental contrasts (i.e., *AL-T0* vs *AL-T2* from Duroc GM muscle and *lean* vs *obese* from UNIK Duroc-Göttingen adipocytes). Besides, from the list of genes with the top highly negative $PT_c$ scores, we only focused on genes showing strong reduced $\Delta Ex$ values > 2 folds for post-transcriptional signals in *AL-T0* vs *AL-T2* animals and reduced $\Delta Ex$ > 3 folds in *lean* vs *obese* animals. The higher fold change in exonic fractions used for the *lean* vs *obese* experimental design was motivated by the overall weaker differential expression obtained at both mRNA and miRNA levels for these data sets. Conversely, genes showing increased $\Delta Int$ values > 2 folds were retrieved as putatively showing transient transcriptional activation. MiRNAs and TFs were considered as the main effectors of the observed target mRNA downregulation and upregulation, respectively.

**Differential expression and significance of regulatory signals**

Differential expression analyses were carried out with the *edgeR* package [33] using the exonic fraction of mRNAs and miRNA expression profiles from RNA-Seq and small RNA-Seq data sets comprising *AL-T0 vs AL-T2* and *lean vs obese contrasts*. Briefly, expression filtered raw counts (CPM > 1 in at least 50% of samples) for exonic reads were normalized with the trimmed mean of M-values normalization (TMM) method [34]. Subsequently, after dispersion estimation with a Cox-Reid profile-adjusted likelihood method [35], a generalized log-linear model of the negative binomial distribution was fitted to the count data and significance in expression differences was tested with a quasi-likelihood F-test implemented in the glmQLFit function from edgeR method [33]. This approach is preferred for experimental designs with small sample number as it provides a more robust and reliable type I error rate control compared with the canonical chi-square approximation to the likelihood ratio test [36]. Correction for multiple hypothesis testing was applied using the Benjamini-Hochberg false discovery rate approach [37]. Genes were considered as differentially expressed (DE) when they had |FC| > 2 and *q*-value < 0.05 for mRNAs, and |FC| > 1.5 and *q*-value < 0.05 for miRNAs.

The statistical significance of the transcriptional ($T_c$) and post-transcriptional ($PT_c$) component variation between groups (*AL-T0 vs AL-T2* and *lean vs obese*) was evaluated using the same edgeR statistical framework with following specifications:

The transcriptional component ($T_c$) significance was based on differences in mean expression at the intronic fraction, which was considered as a prior manifestation of pre-mRNA levels immediately after transcription and before any other further processing in the nucleus. Conversely, the post-transcriptional component ($PT_c$) significance was modeled by incorporating the intronic quantification as an interaction factor for exonic

abundances. In this way we accounted for the previous intronic variation effect on the differences shown at the exonic level for the mature mRNA once in the cytoplasm.

**miRNA target prediction**

Target interactions between miRNA seeds and 3'-UTRs of protein-coding mRNA genes were assessed on the basis of sequence identity using the Sscrofa11.1 reference assembly. Annotated 3'-UTRs from porcine mRNAs were retrieved from Sscrofa11.1 v103 available at bioMart database (http://www.ensembl.org/biomart) and miRBase database [38]. The 3'-UTR sequences shorter than 30 nts were discarded. Redundant seeds from mature porcine microRNAs were removed to avoid redundancy in miRNA target prediction.

The seedVicious v1.1 tool [39] was used to infer miRNA-mRNA interactions against the non-redundant set of porcine miRNA seeds and 3'-UTRs retrieved from the Sscrofa11.1 v103 annotation (http://ftp.ensembl.org/pub/release-103/gtf/sus_scrofa/). miRNA-mRNA interactions of type 8mer, 7mer-m8 and 7mer-A1 were considered as the most relevant and potentially functional among the set of other alternative non-canonical occurring interactions [5, 40, 41]. Following early reports about the effects of miRNA binding site context in determining the miRNA-mRNA interaction efficacy [40], we removed *in silico*-predicted miRNA-mRNA interactions complying with any of the following criteria: (i) Binding sites located in 3'-UTRs at less than 15 nts close to the end of ORF (and the stop codon) or less than 15 nts close to the terminal poly(A) tail. (ii) Binding sites located in the middle of the 3'-UTR in a range comprising 45-55% of the central body of the non-coding sequence. (iii) Binding sites lacking AU-rich elements in their immediate upstream and downstream flanking regions comprising 30 nts each.

**Transcription factor binding site prediction and enrichment analyses**

Target interactions between TFs and promoter regions of mRNA genes were assessed with the *findMotif.pl* script from HOMER software tool [42] using human annotation for promoter regions and default TF binding motif database for vertebrates within HOMER tool [42]. Genes ranked at the top 1% for most negative $PT_c$ scores and having increased ΔInt values > 2 folds were converted to their corresponding human equivalence with the bioMart tool (http://www.ensembl.org/biomart) and used as input for running motif enrichment. We defined a window of 500 nts upstream in the promoter region and 100 nts downstream the transcription start site within the 5'-UTR of each selected gene for TF binding motif search. Analogous analyses were carried out for the list of genes with the top 1% most positive $T_c$ scores and ΔEx values > 2 folds.

The presence of enriched TF motifs within the selected target gene lists was assessed with a hypergeometric test of significance for overrepresentation of binding site motifs in the target promoter sequences. Correction for multiple hypothesis testing was implemented with the Benjamini-Hochberg false discovery rate approach [37]. Enrichment of TF binding motifs were considered significant at *q*-value < 0.05.

**Target-wise and gene-wise miRNA enrichment & depletion analyses**

After the identification of mRNA genes showing marked post-transcriptional signals and prediction of putative miRNA binding sites in their 3'-UTRs, we sought to determine if the overall number of binding sites (target-wise) for upregulated miRNAs (FC > 1.5; *q*-value < 0.05) was significantly enriched. Moreover, we investigated whether the overall number of genes (gene-wise) within those with relevant post-transcriptional signals were also significantly enriched. The whole set of expressed mRNA genes with available 3'-UTRs was used as a contrast background for statistical significance. Enrichment analyses

were carried out using the Fisher's exact test implementation included in the *fisher.test* R function for testing independence of occurrences in a contingency table with fixed marginals. Enrichment results were considered significant at *P*-value < 0.05.

Besides, we also tested whether these genes were enriched at gene-wise and target-wise level for binding sites of the top 5% most highly expressed miRNA genes, excluding significantly upregulated DE miRNAs, as well as for binding sites of significantly downregulated miRNAs (FC < 1.5; *q*-value < 0.05). Upregulated or downregulated miRNAs showing redundant seeds were considered as one single binding site event.

Conversely, we also aimed to investigate whether the set of genes showing marked TF-driven transcriptional activation signals were significantly depleted for being miRNA targets at both gene-wise and target-wise level. Analogously to what was implemented for genes with relevant post-transcriptional signals, we compared the depletion significance against the set of upregulated (FC > 1.5; *q*-value < 0.05) and downregulated (FC < 1.5; *q*-value < 0.05) miRNAs, as well as of the top 5% most highly expressed miRNAs, excluding those significantly upregulated ones. Depletion analyses were carried out using the Fisher's exact test analogously to previously described methods for enrichment analyses considering the opposite distribution tail of occurrences.

Given the relatively low significant differences in miRNA expression obtained for the small RNA-Seq profiles of UNIK Duroc-Gottingen minipigs, miRNAs were considered upregulated at FC >1.5 and *P*-value < 0.01 for this particular experimental setup.

As a control test, we implemented a randomized bootstrap corrected iteration to generate 100 random sets of 10 expressed mature miRNA genes without seed redundancy. These lists were used as input for miRNA target prediction with the sets of prioritized genes with the top 1% and 5% transcriptional and post-transcriptional signals previously described. The distribution of odds ratios obtained after iterating over each random set of

miRNAs (N = 100) were then compared with odds ratios obtained with the set of significantly upregulated miRNAs.

The *P*-value for the significance of the deviation of observed odds ratios against the bootstrapped odds ratios distribution was defined with the following formula:

$$P - value = 1 - \frac{r + 1}{k + 1}$$

where $r$ is the number of permuted odds ratios with values equal or higher than the observed odds ratio for enrichment analyses with the set of upregulated miRNAs, and $k$ is the number of defined permutations (N = 100). For depletion analyses, $r$ was defined as the number of permuted odds ratios with values equal or lower than the observed odds ratio for depletion analyses with the set of upregulated miRNAs.

**Gene covariation network reconstruction**

We hypothesized that genes showing relevant post-transcriptional regulatory effects might be regulated by the same set of significantly upregulated miRNAs, which could induce shared covariation in their expression profiles at the exonic level. On the contrary, their intronic fractions would be unaffected, as introns would have been excised prior to any given miRNA-driven downregulation, if occurring. In this way, an increased gene covariation might be detectable within the sets of commonly targeted mRNA genes at the exon but not at the intron level, as opposed to covariation events of these set of genes with the rest of expressed genes.

In order to test such hypothesis, we computed pairwise correlation coefficients among the whole set of DE mRNA genes (*q*-value < 0.05) in the *AL-T0* vs *AL-T2* experimental

contrast. Pairwise Spearman's correlation coefficients ($\rho$) were calculated for each DE mRNA gene, excluding self-correlation coefficients. Significant covariation events were identified with the Partial Correlation with Information Theory (PCIT) network inference algorithm [43] implemented in the PCIT R package [44]. Using first-order partial correlation coefficients for each trio of genes, along with an information theory approach, this tool is able to identify meaningful informative gene-to-gene covariation once the influence of other spurious covariation in trios has been excluded. Non-significant covarying pairs were set to zero, while significant covarying pairs with both positive or negative coefficients $|\rho| > 0.6$ where assigned a value of 1.

**Covariation enrichment score**

The functional regulatory implications of miRNAs in shaping covariation patterns were then assessed by a covariation enrichment score (CES) calculation following Tarbier et al. 2020 [45]. The CES value for a given gene set reflects the frequency with which the mRNA expression correlation between members of that set is significant relative to the correlation between members of the whole expression data set. In this way, we can represent the variability in covariation within a set of genes as a fold change comparing the increase or reduction of observed significant covariation events relative to the remaining genes in the correlation matrix, and computed as follows:

Being $M$ the pairwise correlation square matrix previously described, i.e., $M = \{m_{ij}\}$, is 1 if the correlation between the pair of genes {i, j} is a significant covarying pair and 0, otherwise. We defined $D = M[i_B, j_B]$, the submatrix corresponding to the set genes of interest, with $i, j_B = \left\{\cup_{index_S}\right\}$, i.e., $D$ is a square matrix of size $d$ , being $d$ the number of gene set of interest (those with the top 1% (N = 12) and 5% (N = 18) post-transcriptional signals and putatively targeted by upregulated DE miRNAs). The, the

enrichment of the given subset is computed with the following three steps: (i) For each row ( $i = 1..d$ ) of $D$ , compute $s_i. = \sum_{l=1}^{k} \frac{d_{ik}}{d}$ , i.e., $i \in i_B$ (ii) Being the matrix $M[i_B,.]$, which considers only the rows of $M$ corresponding to the genes of interest and all of its columns ($p$), compute $t_i. = \sum_{l=1}^{p} \frac{m_{il}}{p}$ $\forall i \in i_B$, (iii) $s_i.$ and $t_i.$ are vectors of size $d$, then, the CES value for each gene $i \in i_B$ is computed as the ratio between $s$ and $t$, $CES_i = \frac{s_i}{t_i}$.

Additionally, we defined random sets of genes of same length as those previously defined for genes with post-transcriptional signals and the process was repeated iteratively (N = 1,000) over the set of differentially expressed genes ($q$-value < 0.05) in the *AL-T0* vs *AL-T2* contrast. We computed the CES values for all resulting random sets of genes and they were considered as a control with no expected enrichment in covariation patterns.

Differences among the set of exonic, intronic and control CES values were tested with a non-parametric approach using a Mann-Whitney U test [46].

**Expression analyses of miRNAs and putative mRNA targets by qPCR**

Retroperitoneal adipose tissue (~20 ml) was taken from the abdominal cavity of UNIK intercrossed Duroc-Göttingen minipigs pigs quickly after slaughtering (more details about UNIK minipig population are described elsewhere [23, 24]). Adipocytes were isolated as described in Jacobsen et al. 2019 [20] and used for RNA isolation following the method of Cirera (2013) [26]. Total RNA from adipocytes used for RNA-Seq and small RNA-Seq profiling as previously described was subsequently employed for quantitative real-time polymerase chain reaction (qPCR) verification of expression changes. Five mRNAs (*LEP*, *OSBLP10*, *PRSS23*, *RNF157* and *SERPINE2*) among the top 5% with the highest negative $PT_c$ values and showing ΔEx higher than 3-fold

(equivalent to 1.58 $\log_2$-folds) were selected for expression profiling. The same 5 *lean* and 5 *obese* animals (according to BMI profiles, **Table S1**) used for RNA-Seq and small RNA-seq analyses [20, 47] were selected, with the exception of animal 572, for which no additional RNA was available and that was replaced by animal 503, which had the closest BMI profile within the *obese* group of animals (**Table S1**).

For mRNA profiling, 400 ng of total RNA were used for cDNA synthesis in duplicate. Briefly, 0.5 µl Improm-IITM reverse transcriptase (Promega), 0.25 µg 3:1 mixture of random hexamers/OligodT (Sigma-Aldrich), 2 µl 5× ImProm-II buffer, 10 units RNasin Ribonuclease inhibitor (Promega), 2.5 mM $MgCl_2$ and 2mM dNTPs were mixed with total RNA in a final reaction volume of 10 µl following manufacturer's instructions. Each reaction was incubated during 5 min at room temperature, 1 h at 42°C and 15 min at 70°C for enzyme inactivation. A negative control (RT) without reverse transcriptase reaction was included. Finally, the synthesized cDNAs were diluted in a 1:16 proportion and subsequently used for qPCR analyses. Two reference genes (*TBP* and *ACTB*, according to Nygard et al. 2007 [48]) were used as normalizers. Primers were collected from available stocks from previous studies [48, 49] or designed using the PRIMER3 software within the PRIMER-BLAST [50], considering exon–exon junction spanning primers for poly-exonic candidates and accounting for multiple transcript capture when possible.

Accordingly, three miRNAs among the most significantly differentially expressed in small RNA sequencing data from UNIK samples were selected for qPCR profiling (ssc-miR-92b-3p, ssc-miR-148a-3p and ssc-miR-214-3p) plus two miRNAs for normalization which were among the most highly expressed and with no differential expression signal in the small RNA-Seq data set (ssc-let-7a and ssc-miR-23a-3p). The same RNA samples used for mRNA profiling were subsequently processed for microRNA profiling using 50 ng of total RNA for cDNA synthesis in duplicate and following the protocol for

microRNAs of Balcells et al. 2011 [51] with recommendations reported by Cirera & Busk 2014 [52] and synthesized cDNAs were diluted in a 1:8 proportion before use in qPCR reaction. Primers for miRNA qPCR amplification were designed using the miRprimer software [53]. All primers for mRNA and miRNA expression profiling are available at **Table S2.**

QPCR for mRNA and miRNA was performed on a MX3005P machine (Stratagene, USA) using the QuantiFast Multiplex PCR protocol kit (Qiagen, Germany). Briefly, diluted cDNAs were mixed with 5µl of 2× QuantiFast Multiplex master mix (Qiagen, Germany) and 10µM of each primer (**Table S2**) were mixed in a final volume of 10µl. Cycling conditions were: 95 °C for 5 min followed by 40 cycles of 95 °C for 10 sec and 60 °C for 30 sec. Melting curve analyses (60 °C to 99 °C) were performed after completing amplification reaction to ensure the specificity of the assays. Data were processed with the MxPro qPCR associated software. Assays were considered successful when: 1) the melting curve was specific (1 single peak) and 2) the samples had Cq values < 33 cycles (i.e., sufficiently expressed to be considered biologically functional).

Efficiency of each primer assay was estimated from the log–linear fraction of their corresponding standard curves. Primer efficiency ranging between 80–110% was considered successful with $R^2 > 0.98$. Pre-processing of raw Quantification cycle (Cq) values was performed with the GenEx v.6 software (MultiD, Analyses AB). Briefly, raw expression data for each successfully amplified gene was corrected according to its corresponding PCR efficiency and Cq values were normalized using mRNA and miRNA normalizers previously evaluated with the *geNorm* algorithm [54], respectively. Duplicates for each sample were subsequently averaged and relative quantities were scaled to the lowest expressed sample in each assay and $\log_2$ transformed before statistical analyses. Raw Cq values for each assay are available at **Table S3**. The significance of

differential expression between *lean* and *obese* animals was estimated with the limma empirical Bayes procedure from *limma* R package [55] incorporating a mean-variance trend modeling the relationship between variance and gene signal intensity and correcting for sex as fixed covariate when fitting the linear mixed model prior to the use of empirical *ebayes* function for analyzing difference between the mean expression in *lean* vs *obese* animals.

# Results

**Determining transcriptional and post-transcriptional signals with EISA approach**

After the processing, mapping and quantification of mRNA and miRNA gene expression profiles in GM skeletal muscle samples encompassing 11 fasting (*AL-T0*) and 12 fed (*AL-T2*) Duroc gilts, an average of 45.2 million reads per sample (~93%) successfully mapped to annotated genes (N = 31,908) in the Sscrofa11.1 assembly. Besides, around 2.2 million reads per sample (~42%) from an average of 6.2 million small RNA sequences mapped to annotated porcine miRNA genes (N = 370).

A total of 30,322 genes based on exonic regions and 22,769 genes based on intronic regions were identified after splitting the reference Sscrofa11.1 v.103 assembly between exonic and intronic ranges. The reduced amount of genes for intronic ranges was produced upon the removal of singleton regions, mono-exonic genes and exon-overlapping intronic ranges that could disturb the accurate determination of intronic fractions.

The exonic fraction accounted for an average of 1,923.94 estimated counts per gene, whereas the intronic fraction was represented by an average of 83.02 counts per gene, meaning that counts in exonic ranges exceeded those in intronic ranges by ~23 fold.

EISA analyses were run using exonic and intronic fractions from the *AL-T0* vs *AL-T2* contrast excluding genes with expression levels below 1 CPM in at least 50% of the samples (N = 12). Only genes represented by both exonic and intronic mapping counts were retained, resulting in a total of 9,492 mRNA genes that were subsequently used for determining ΔEx and ΔInt values, using the fasting group (*AL-T0*) as baseline control. This represents any given upregulation in ΔEx or ΔInt values as an overexpression in fed (*AL-T2*) over fasting (*AL-T0*) Duroc gilts. $T_c$ and $PT_c$ scores were obtained from z-scored values of ΔInt and ΔEx – ΔInt estimates, respectively. The statistical significance of observed transcriptional ($T_c$) and post-transcriptional ($PT_c$) signals were obtained as described in Methods section. Moreover, differential expression analyses based on exonic fractions were carried out using the 9,942 genes retrieved previously.

Differential expression analyses resulted in a total of 454 DE genes (*q*-value < 0.05), and of those, only genes with |FC| > 2 were retained, resulting in 52 upregulated and 80 downregulated genes (**Table S4, Fig. S1a**). Besides, differential expression analyses on small RNA-seq data for *AL-T0* vs *AL-T2* pigs revealed a total of 16 DE miRNAs (|FC| > 1.5; *q*-value < 0.05), of which 8 were downregulated and the remaining ones were upregulated (**Table S5**). The non-redundant seeds of miRNAs with significantly differential upregulation in fed *AL-T2* animals (N = 6; ssc-miR-148a-3p, ssc-miR-7-5p, ssc-miR-30-3p, ssc-miR-151-3p, ssc-miR-374a-3p and ssc-miR-421-5p, **Table S5**) were selected as those potentially affecting genes post-transcriptionally after nutrient supply. $PT_c$ signal analyses revealed a total of 133 genes with significant post-transcriptional components (|FC| > 2; *q*-value < 0.05, **Table S6**), of which three had reduced ΔEx fractions > 2 folds and two of them had negative $PT_c$ scores suggestive of post-transcriptionally miRNA-driven downregulation. Regarding $T_c$ signal analyses, a total of 447 genes had significant transcriptional components (|FC| > 2; *q*-value < 0.05, **Table**

**S6**), of which 164 were upregulated and 156 were downregulated with |ΔInt| fractions > 2 folds.

From these results, we aimed at determining relevant genes putatively downregulated by miRNAs at the post-transcriptional level, as well as genes transiently upregulated by active transcription activation mediated by TF activity. In this way, we established two filtering criteria in order to prioritize candidate genes for both effects, separately. First, genes within the top most extreme 1% and 5% negative $PT_c$ scores with visibly reduced ΔEx values > 2 folds were selected as genes showing putative miRNA-driven post-transcriptional downregulation (**Fig. S1b**). This resulted in a total of 13 and 26 selected genes, respectively (**Table 1**), from which one of them did not have a 3´-UTR annotated (ENSSSCG00000049158) and was hence excluded from miRNA seed target prediction at the 3'-UTR level, giving a total of 12 and 25 genes, respectively. Among the genes with the top detected post-transcriptional signals were, to mention a few, the Dickkopf WNT Signaling Pathway Inhibitor 2 (*DKK2*), pyruvate dehydrogenase kinase 4 (*PDK4*), interleukin 18 (*IL18*), Nuclear receptor subfamily 4 group A member 3 (*NR4A3*), Cholinergic receptor nicotinic α1 subunit (*CHRNA1*) or myosin IXA (*MYO9A*).

Such approach was motivated by the notion that genes showing strong downregulation in exonic variance (ΔEx) between groups, combined with highly negative $PT_c$ scores (ΔEx – ΔInt), would represent a proxy for post-transcriptionally regulated genes. The strong downregulation observed in ΔEx values would exceed the observed ΔInt fraction, resulting in a lack of intronic downregulation that might result from post-transcriptional changes only at the exonic level after mRNA splicing in the nucleus.

We did not consider the significance of $PT_c$ scores as a relevant criterion for prioritizing putative post-transcriptionally downregulated genes, as these will appear as significant only when the post-transcriptional activity was the only mechanism modulating the target

gene expression profile. Because coordinated regulatory effects at both transcriptional and post-transcriptional levels might occur, these events would result in non-significant $PT_c$ scores given the confounding effect introduced in statistical analyses. Only co-occurring yet opposite transcriptional and post-transcriptional events or isolated post-transcriptional signals would arise as significant, excluding those genes with both coordinated up or downregulation at the post-transcriptional level.

Conversely, for genes showing putative active TF-driven transcription activation, those with the top most extreme 1% and 5% negative $PT_c$ scores and with ΔInt values > 2 folds were selected (**Fig. S1c**), resulting in a total of 53 and 117 mRNA genes, respectively (**Table S7**). Among these genes were, to mention a few, the kyphoscoliosis peptidase (*KY*), heat shock protein 10 (*HSPE1*), heat shock protein family B member 1 (*HSPB1*), carnosine synthase 1 (*CARNS1*), fibronectin type III and SPRY domain containing 2 (*FSD2*), serpin family member H 1 (*SERPINH1*), anoctamin 1 (*ANO1*), calpain 1 (*CAPN1*) or striated muscle enriched protein kinase (*SPEG*).

Genes showing strong upregulation at the intronic (ΔInt) but not in the exonic (ΔEx) level would also result in highly negative $PT_c$ scores (ΔEx – ΔInt) and could be considered as a proxy of active transient transcriptional activity within the nucleus. The observed excess in the intronic fraction might reflect an active transcription of yet unspliced pre-mRNA transcripts leading to the accumulation of intronic sequences prior to their debranching and degradation by exonucleases. On the contrary, genes with coordinated upregulation of both exonic and intronic fractions would result in elevated and significant $T_c$ scores and $PT_c$ scores close to zero, possibly reflecting previous transcription activation events yet still detectable at the transcriptomic level.

**Context-based pruning of predicted miRNA-mRNA interactions removes spurious unreliable target events**

As a first step to determine if genes with highly negative $PT_c$ scores and showing a marked reduction in exonic fractions were possibly affected by the repressor activity of miRNAs, we aimed to investigate the accuracy and reliability of *in silico* predicted miRNA binding sites in their 3'-UTRs (**Table S8**).

We evaluated the presence of enriched binding sites gene-wise over a random background without any context-based filtering, applying each one of the three established criteria independently (see methods), and combining them pairwise and altogether (i.e., we analyzed how many of the genes were putative targets of upregulated DE miRNAs in the *AL-T0* vs *AL-T2* contrast with and without additional filters on putative miRNA binding site prediction). As depicted in **Fig. 1e-f**, introducing additional context-based filtering criteria for removing spurious unreliable binding site predictions resulted in an overall increased enrichment of miRNA targeted genes within the top 1% (**Fig. 1e**) and 5% (**Fig. 1f**) negative $PT_c$ scores and with reduced $\Delta Ex > 2$ folds. This increase in the enrichment significance was more prevalent for the AU-rich-based criterion, and was slightly improved when adding the other two context-based filtering criteria (**Fig. 1e**). These results were, however, less evident for the list of the top 5% genes (**Fig. 1f**), probably due to the reduced stringency of gene prioritization and the inclusion of putative false positive candidate genes. Nevertheless, still an increased enrichment was detectable for all combined filtering criteria, especially for binding sites of type 7mer-A1, and probably at the expense of the scarcer and more efficient 8mer binding sites.

Supported by the observed increased reliability of retrieved miRNA binding sites and targeted genes after context-based site pruning, we applied the three joint aforementioned

criteria for further enrichment/depletion analyses of miRNA-derived post-transcriptional regulatory signals.

**Genes with relevant post-transcriptional signals are enriched for putative miRNA binding sites in their 3'-UTRs.**

Target prediction and context-based pruning of miRNA-mRNA interactions for genes with the top 1% negative $PT_c$ scores and a reduction in exonic fraction of at least 2 folds (N = 12) resulted in a total of 8 binding sites of type 8mer, 13 of type 7mer-m8 and 9 of type 7mer-A1 (**Table S8**). These binding sites belonged to non-redundant seeds of DE upregulated miRNAs in *AL-T2* gilts (ssc-miR-148a-3p, ssc-miR-7-5p, ssc-miR-30-3p, ssc-miR-151-3p, ssc-miR-374a-3p and ssc-miR-421-5p, **Table S5**). Besides, the number of binding sites within the 3´-UTRs from genes with the top 5% negative $PT_c$ scores and a reduction in exonic fraction of at least 2 folds (N = 25) were 11 of type 8mer, 21 of type 7mer-m8 and 22 of type 7mer-A1 (**Table S8**). When we compared these with the number of predicted binding sites of all expressed genes as background, including those previously analyzed (N = 9,492), we observed a significant enrichment for miRNA binding sites present in the 3'-UTRs of the 12 genes with putative post-transcriptional signals showing the top 1% $PT_c$ scores (**Fig. 2a**). This enrichment was more prevalent in binding sites of type 8mer + 7mer-m8 combined, followed by 8mer targets alone.

Moreover, we also evaluated the enrichment of the number of binding sites for these two set of genes with the following sets of miRNAs: (i) Non-redundant seeds from downregulated miRNAs in *AL-T2* fed gilts (ssc-miR-1285, ssc-miR-758, ssc-miR-339, sc-miR-22-3p, ssc-miR-296-5p, ssc-miR-129a-3p, ssc-miR-181c and ssc-miR-19b, **Table S5**), (ii) the top 5% most expressed non-redundant miRNA seeds, excluding those being upregulated, if present (ssc-miR-1, ssc-miR-133a-3p, ssc-miR-26a, ssc-miR-10b,

ssc-miR-378, ssc-miR-99a-5p, ssc-miR-27b-3p, ssc-miR-30d, ssc-miR-486 and ssc-let-7f-5p) and (iii) for an iteration (N = 100) of random sets of 10 expressed miRNAs, irrespective of their DE and abundance status, as a control test. None of these additional analyses recovered a significant enrichment for any type of the three considered miRNA target subtypes (**Fig. 2a**).

Regarding genes with the top 5% $PT_c$ scores (N = 25, **Fig. 2b**), enrichment analyses did not reach significant results for the number of binding sites in their 3'-UTRs, but they overall followed similar trends compared with analyses focused on the top 1% of $PT_c$ scores previously described. However, the predicted miRNA binding sites were still more abundant than those obtained for enrichment analyses considering downregulated or highly expressed miRNAs (**Fig. 2b**).

To exclude the possibility that additional DE downregulated genes (**Table S4**) were also putatively regulated by miRNAs, removing those previously analyzed for post-transcriptional signals (which were not necessarily downregulated in a significant manner), we also repeated our enrichment analyses for binding sites on this set of genes (N = 48). Our results showed that these mRNA genes gathered nearly no binding sites from the set of upregulated DE miRNA genes (**Table S5**), and this observation was also reproduced for downregulated DE miRNAs, as well as for the top 5% highly expressed miRNAs, although a more prevalent number of binding sites were obtained for the latter (**Fig 2c**). On the contrary, the average odds ratios for enrichment analyses of binding sites obtained for random sets of miRNA seeds were predominantly higher (i.e., a more significant enrichment) than those observed for targets of upregulated miRNAs, thus indicating the overall depletion in binding sites for this set of mRNAs with no sign of post-transcriptional regulatory effects, yet still significantly downregulated (**Fig. 2c**).

Additionally, we also aimed at investigating if the genes showing putative post-transcriptional signals were enriched for miRNA-derived target events at the gene-wide level (i.e., considering the number of overall genes being targeted, instead of the number of binding sites as previously described). Both sets of genes with the top 1% (N = 12, **Fig. 3a**) and 5% (N = 25, **Fig. 3c**) $PT_c$ scores and reduced ΔEx > 2 folds obtained significant enrichment results for 8mer, 7mer-m8 and 7mer-A1 targeted genes (**Fig. 3b** and **3d**), and this was especially relevant for all target types combined. More importantly, all 12 genes from the top 1% and 21 out of 25 genes from the top 5% $PT_c$ scores (**Table S8**) were predicted as putative targets for miRNAs upregulated in *AL-T2* fed gilts (**Table S5**).

A special case was represented by the ENSSSCG00000049158 gene, which had a $PT_c$ score among the top 1% with a reduced ΔEx > 2 folds, but with no 3'-UTR available in the Sscrofa11.1 annotation, which led to its removal for miRNA-mRNA target prediction analyses. On the contrary, this gene has a relatively long and fragmented 5'-UTR, and we decided to investigate whether putative miRNA targets could happen in this region. No context-based filtering was implemented for these analyses. Surprisingly, we found a total of 7 and 9 binding sites (**Table S9**) for the set of DE upregulated miRNAs (**Table S5**) across the two annotated transcripts of this gene, respectively.

Similar to what we obtained when using the set of downregulated miRNAs or those being highly expressed, no significant enrichment was observed, except for a slight yet relevant enrichment present for miRNA-mRNA interactions between genes with post-transcriptional regulatory signals and highly expressed miRNAs (**Fig. 3b**). When we analyzed the enrichment of the number of DE downregulated genes (**Fig. 3e**) being putative targets of upregulated, downregulated or highly expressed miRNAs, no significant results were observed, except for genes targeted for highly expressed miRNAs considering 8mer + 7mer-m8 binding sites only (**Fig. 3f**).

**Genes showing post-transcriptional regulatory signals predominantly covary at the exonic level.**

To further investigate whether genes with relevant post-transcriptional regulatory signals could account for truly targeted genes by miRNAs, we evaluated the covariation patterns among them and with the whole remaining DE mRNA genes (q-value < 0.05, N = 454) in our RNA-seq data from *AL-T0* and *AL-T2* gilts. In this way, if a set of genes were to be downregulated by any upregulated miRNAs in a coordinated manner, we would expect a reduced abundance in their mature spliced mRNA forms, i.e., only the exonic fractions but not the intronic fractions might predominantly covary. The intronic fraction, eventually spliced and degraded in the nucleus, should not reflect any posterior post-transcriptional regulatory effects in the cytoplasm, and reduced or null covariation in their abundances might be expected.

Using CES values previously described, we obtained an estimation of the fold change in the observed covariation of our set of putatively miRNA-targeted genes with respect to other DE mRNAs, and this was measured for both their exonic and their intronic abundances. As depicted in **Fig. 4**, both sets of miRNA target genes with the top 1% (N = 12, **Fig. 4a**) and 5% (N = 18, **Fig. 4b**) $PT_c$ scores and with reduced $\Delta Ex > 2$ folds showed average significantly increased covariation rates of ~1.5 and ~2 folds in their exonic fractions compared with their intronic fractions, respectively. Moreover, when we analyzed the observed fold change in covariation for random sets of genes iteratively (N = 1,000), they fell towards null covariation changes, i.e., CES ≈ 1 (**Fig. 4**). The observed CES distributions of exonic, intronic and control sets were significantly different after running non-parametric tests (**Fig. 4**), thus supporting the expected predominant effect at the exon level by miRNA-driven post-transcriptional downregulation.

**Genes with relevant transcriptional signals are depleted for putative miRNA binding sites in their 3'-UTRs.**

Analogously to what we implemented for enrichment analyses on genes with post-transcriptional signals and miRNA-derived binding sites, we then investigated whether genes with the top 1% (N = 53, **Fig. 5a**) and 5% (N = 117, **Fig. 5c**) $T_c$ scores and ΔInt > 2 folds (**Table S7**) had signals of transcriptional transient activation, which would mean the absence of, in principle, any miRNA-derived post-transcriptional regulation, as pre-mRNAs would be still accumulating in the nucleus prior to any further post-transcriptional effects once in the cytoplasm. This effect would be seen as an excess of the observed intronic fraction compared with the exonic fraction, leading to negative $PT_c$ scores (ΔEx – ΔInt). After miRNA target prediction, a total of 18 out of 53 genes (top 1% negative $PT_c$ and increased ΔInt > 2 folds) were targeted by DE upregulated miRNAs in fed AL-T2 gilts (**Table S5**), whereas 42 out of 117 genes were targeted when considering the top 5% negative $PT_c$ and increased ΔInt > 2 folds (**Table S10**).

In this regard, depletion analyses for gene-wise targeted mRNAs by upregulated miRNAs (**Table S5**) were implemented, resulting in an observed significant depletion for all considered target types (**Fig. 5b** and **Fig. 5d**), and this was especially relevant for genes with the top 5% negative $PT_c$ scores and ΔInt > 2 folds (**Fig. 5d**, **Table S10**). Contrary to previous observations, mRNA genes with transcriptional activation signals were also significantly depleted for being targeted by downregulated or highly expressed miRNAs (**Fig. 5b** and **Fig. 5d**), thus reinforcing the notion that these transcripts might be indeed under active transcription in the nucleus prior to any further miRNA-derived post-transcriptional regulation in the cytoplasm.

Finally, after running depletion analyses using the set of DE upregulated genes (**Fig. 5e**, **Table S4**), and excluding those with active transcriptional upregulation signals (**Table S7**), a significant depletion in the number of targeted mRNA genes by upregulated miRNAs (**Table S5**) was observed (**Fig. 5f**), but this was not reproduced for downregulated or highly expressed miRNAs. These results indicated that DE upregulated genes (**Fig. 5e**) other than those with active transient transcriptional signals (**Fig. 5a** and **5c**) might be subjected to a certain post-transcriptional effect driven by highly expressed miRNAs in the cytoplasm.

**TF-binding motif enrichment analyses predict genes with active transient transcriptional upregulation.**

From the list of genes with putative transient transcriptional activation, we wanted to investigate whether they might share transcription factor binding sites that could explain their coordinated transcription response to feeding intake in *AL-T2* gilts. By searching for TF binding sites on their promoter regions, we identified a nearly enriched motif (**Fig. S2,** *q*-value = 6.54E-02) targeted by the myocyte enhancer factor 2D (*MEF2D*) transcription factor, which was predicted to bind the promoter region of 7 out of 53 genes (**Table S11**) from those with the top 1% negative $PT_c$ scores and ΔInt > 2 folds (**Table S7**). This TF binding motif showed a significant enrichment (*q*-value < 0.1, **Fig. S2**) compared with a background set of promoter regions by using the *findMotif* algorithm from HOMER software tool. Moreover, we wanted to test whether other additional DE upregulated genes might also share enriched TF binding motifs. To do so, we took the list of upregulated DE genes (N = 52) and also those showing the top 1% $T_c$ scores and ΔEx > 2 folds, while excluding any genes overlapping with those within the top negative $PT_c$ scores with transient transcriptional activation signals (**Table S7**). Although a number of

shared TF binding motifs were identified (**Fig. S2**), none of them were significantly enriched as opposed to the *MEF2D* binding site for genes with relevant transcriptional signals.

**Studying post-transcriptional signals in adipocytes metabolism using an independent Duroc-Göttingen minipig population**

We evaluated the performance of EISA analyses on the miRNA target prioritization using the adipocyte mRNA and miRNA expression profiles of an additional Duroc-Göttingen minipig population with divergent fatness profiles well characterized at the metabolic and molecular level [23, 24]. A total of 10 animals divided into two groups of 5 animals each with high and low BMI profiles (*lean* vs *obese*, **Table S1**) were selected for sequencing of adipocyte homogenates collected from retroperitoneal adipose tissue. After pre-processing and filtering of sequenced reads, an average of ~98.1 and ~0.87 million mRNA and small RNA reads per sample were generated, of which ~96.5% and ~73.4% mapped to annotated porcine mRNA and miRNA genes, respectively. Differential expression analyses revealed a total of 299 DE mRNAs ($q$-value < 0.05), of which 52 mRNAs were downregulated and 95 were upregulated (|FC| > 2; $q$-value < 0.05), respectively (**Table S12**). Regarding miRNAs, only one gene (ssc-miR-92b-3p) was significantly upregulated in *lean* pigs (|FC| > 2; $q$-value < 0.05), while 7 additional miRNAs showed suggestive differential expression ($P$-value < 0.01, **Table 2**). The whole set of miRNA genes from differential expression analyses is available at **Table S13**.

After running EISA analyses on the mRNA expression profiles for exonic and intronic fractions, a total of 15 downregulated mRNAs in *lean* pigs were among the top 5% $PT_c$ scores with reduced ΔEx > 3 folds, respectively (**Table 3, Fig. 6a**). The whole set of mRNA genes from $PT_c$ EISA analyses is available at **Table S14**. A total of 12 mRNA

genes were then classified as putative miRNA targets from the set of upregulated miRNA genes in *lean* pigs (ssc-miR-92b-3p, ssc-miR-148a-3p, ssc-miR-204 and ssc-miR-214-3p; **Table 2**), respectively. Among these, it is worth mentioning leptin (*LEP*), oxysterol binding protein like 10 (*OSBPL10*), serine protease 23 (*PRSS23*), ring finger protein 157 (*RNF157*), serpin family E member 2 (SERPINE2), estrogen related receptor γ (*ESRRG*), prostaglandin F receptor (*PTGFR*) or cadherin 1 (*CDH1*).

Gene-wise enrichment analyses for the set of putative miRNA target genes with post-transcriptional signals (N = 12, **Table S15**) revealed a slight yet consistent enrichment for miRNA targets of type 7mer-m8 and 7mer-A1 (**Fig. 6b**). Similar to previously described results, then we assessed the presence of enriched miRNA target genes with the set of downregulated miRNAs in *lean* pigs (ssc-miR-190a and ssc-miR-1839-5p), as well as those within the top 5% most highly expressed miRNAs (**Table S13**) and, again, no significant results were obtained (**Fig. 6b**).

Furthermore, in order to validate the results obtained for mRNAs with high post-transcriptional signals and the miRNAs putatively interacting with them, we selected 5 mRNAs (*LEP*, *OSBPL10*, *PRSS23*, *RNF157* and *SERPINE2*) and 3 miRNAs (ssc-miR-148a-30, ssc-miR-214-3p and ssc-miR-92b-3p) for qPCR expression profiling verification.

All the analyzed mRNA genes showed a reduced expression in *lean* pigs compared with their *obese* counterparts (**Fig. 6c**) and the *LEP* gene was the most significantly downregulated gene (*P*-value = 1.12e-10), a result that was in agreement with the strong downregulation observed in differential expression analyses of the RNA-Seq data set (**Table S12**). With regard to miRNAs, the oppositive expression pattern was revealed, with all the three profiled miRNA genes being upregulated in *lean* pigs. Moreover, as described in the small RNA-Seq differential expression results (**Table 2** and **S13**), the

ssc-miR-92b-3p gene was the miRNA with the most significant upregulation observed in qPCR analyses (*P*-value = 3.57e-02, **Fig. 6d**).

# References


1. Martinez NJ, Walhout AJM. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. BioEssays. 2009;31:435–45.

2. Spitz F, Furlong EEM. Transcription factors: From enhancer binding to developmental control. Nature Reviews Genetics. 2012;13:613–26.

3. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173:20–51. doi:10.1016/j.cell.2018.03.006.

4. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172:650–65.

5. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173:20–51. doi:10.1016/J.CELL.2018.03.006.

6. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008;455:64–71.

7. Ebert MS, Sharp PA. Roles for MicroRNAs in conferring robustness to biological processes. Cell. 2012;149:515–24. doi:10.1016/j.cell.2012.04.005.

8. Gaidatzis D, Burger L, Florescu M, Stadler MB. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat Biotechnol. 2015;33:722–9.

9. Cursons J, Pillman KA, Scheer KG, Gregory PA, Foroutan M, Hediyeh-Zadeh S, et al. Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT. Cell Syst. 2018;7:77-91.e7.

10. Pillman KA, Scheer KG, Hackett-Jones E, Saunders K, Bert AG, Toubia J, et al. Extensive transcriptional responses are co-ordinated by microRNAs as revealed by Exon–Intron Split Analysis (EISA). Nucleic Acids Res. 2019.

11. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (80- ). 2008;322:1845–8.

12. Khodor YL, Rodriguez J, Abruzzi KC, Tang C-HA, Marr MT, Rosbash M. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev. 2011;25:2502–12. doi:10.1101/gad.178962.111.

13. Zeisel A, Köstler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, et al. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. Mol Syst Biol. 2011;7:529. doi:10.1038/msb.2011.62.

14. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560:494–8. doi:10.1038/s41586-018-0414-6.

15. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol. 2011;18:1435–40. doi:10.1038/nsmb.2143.

16. Gray JM, Harmin DA, Boswell SA, Cloonan N, Mullen TE, Ling JJ, et al. SnapShot-Seq: a method for extracting genome-wide, in vivo mRNA dynamics from a single total RNA sample. PLoS One. 2014;9:e89673. doi:10.1371/journal.pone.0089673.

17. Hendriks G-J, Gaidatzis D, Aeschimann F, Großhans H. Extensive oscillatory gene expression during C. elegans larval development. Mol Cell. 2014;53:380–92. doi:10.1016/j.molcel.2013.12.013.

18. Zhang D, Tian J, Roden C, Lu J. Post-transcriptional Regulation is the Major Driver of microRNA Expression Variation. bioRxiv. 2020;:2020.01.07.897975.

19. Cardoso TF, Quintanilla R, Tibau J, Gil M, Mármol-Sánchez E, González-Rodríguez O, et al. Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle. BMC Genomics. 2017;18:603. doi:10.1186/s12864-017-3986-x.

20. Jacobsen MJ, Havgaard JH, Anthon C, Mentzel CMJ, Cirera S, Krogh PM, et al. Epigenetic and Transcriptomic Characterization of Pure Adipocyte Fractions From Obese Pigs Identifies Candidate Pathways Controlling Metabolism. Front Genet. 2019;10. doi:10.3389/fgene.2019.01268.

21. Mármol-Sánchez E, Ramayo-Caldas Y, Quintanilla R, Cardoso TF, González-Prendes R, Tibau J, et al. Co-expression network analysis predicts a key role of microRNAs in the adaptation of the porcine skeletal muscle to nutrient supply. J Anim Sci Biotechnol. 2020;11:10. doi:10.1186/s40104-019-0412-z.

22. Ballester M, Amills M, González-Rodríguez O, Cardoso TF, Pascual M, González-Prendes R, et al. Role of AMPK signalling pathway during compensatory growth in pigs. BMC Genomics. 2018;19:682. doi:10.1186/s12864-018-5071-5.

23. Kogelman LJA, Kadarmideen HN, Mark T, Karlskov-Mortensen P, Bruun CS, Cirera S, et al. An F2 pig resource population as a model for genetic studies of obesity and obesity-related diseases in humans: Design and genetic parameters. Front Genet. 2013;4 MAR.

24. Pant SD, Karlskov-Mortensen P, Jacobsen MJ, Cirera S, Kogelman LJA, Bruun CS, et al. Comparative analyses of QTLs influencing obesity and metabolic phenotypes in pigs and humans. PLoS One. 2015;10.

25. Decaunes P, Estève D, Zakaroff-Girard A, Sengenès C, Galitzky J, Bouloumié A. Adipose-derived stromal cells: cytokine expression and immune cell contaminants.

Methods Mol Biol. 2011;702:151–61. doi:10.1007/978-1-61737-960-4_12.

26. Cirera S. Highly efficient method for isolation of total RNA from adipose tissue. BMC Res Notes. 2013;6:472. doi:10.1186/1756-0500-6-472.

27. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10. doi:10.14806/ej.17.1.200.

28. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60. doi:10.1038/nmeth.3317.

29. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.

30. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol. 2013;9:e1003118. doi:10.1371/journal.pcbi.1003118.

31. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47:e47–e47. doi:10.1093/nar/gkz114.

32. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. doi:10.1093/bioinformatics/btt656.

33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40. doi:10.1093/bioinformatics/btp616.

34. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25. doi:10.1186/gb-2010-11-3-r25.

35. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40:4288–97. doi:10.1093/nar/gks042.

36. Lun ATL, Chen Y, Smyth GK. It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. In: Methods in Molecular Biology. Humana Press Inc.; 2016. p. 391–416. doi:10.1007/978-1-4939-3578-9_19.

37. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995;57:289–300. doi:10.2307/2346101.

38. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: From microRNA sequences to function. Nucleic Acids Res. 2019;47:D155–62.

39. Marco A. SeedVicious: Analysis of microRNA target and near-target sites. PLoS One. 2018;13.

40. Grimson A, Farh KKH, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Mol Cell. 2007;27:91–105.

41. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19:92–105.

42. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010;38:576–89. doi:10.1016/j.molcel.2010.05.004.

43. Reverter A, Chan EKF. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics.

2008;24:2491–7. doi:10.1093/bioinformatics/btn482.

44. Watson-Haigh NS, Kadarmideen HN, Reverter A. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. Bioinformatics. 2010;26:411–3. doi:10.1093/bioinformatics/btp674.

45. Tarbier M, Mackowiak SD, Frade J, Catuara-Solarz S, Biryukova I, Gelali E, et al. Nuclear gene proximity and protein interactions shape transcript covariations in mammalian single cells. Nat Commun. 2020;11. doi:10.1038/s41467-020-19011-5.

46. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Stat. 1947;18:50–60. doi:10.1214/aoms/1177730491.

47. Sørensen PM. MicroRNA's in obesity and obesity related phenotypes: Using pig as a model. University of Copenhagen; 2014.

48. Nygard AB, Jørgensen CB, Cirera S, Fredholm M. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Mol Biol. 2007;8:67.

49. Mentzel CMJ, Alkan F, Keinicke H, Jacobsen MJ, Gorodkin J, Fredholm M, et al. Joint Profiling of miRNAs and mRNAs Reveals miRNA Mediated Gene Regulation in the Göttingen Minipig Obesity Model. PLoS One. 2016;11:e0167285. doi:10.1371/journal.pone.0167285.

50. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115–e115. doi:10.1093/nar/gks596.

51. Balcells I, Cirera S, Busk PK. Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. BMC Biotechnol. 2011;11:70. doi:10.1186/1472-6750-11-70.

52. Cirera S, Busk PK. Quantification of miRNAs by a Simple and Specific qPCR Method. In: Methods in molecular biology (Clifton, N.J.). 2014. p. 73–81. doi:10.1007/978-1-4939-1062-5_7.

53. Busk PK. A tool for design of primers for microRNA-specific quantitative RT-qPCR. BMC Bioinformatics. 2014;15:29. doi:10.1186/1471-2105-15-29.

54. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:research0034.1. doi:10.1186/gb-2002-3-7-research0034.

55. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

**Table 1:** Genes with the top 1% and 5% post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds (equivalent to -1 in the $log_2$ scale) from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

| mRNA | Gene | $log_2FC^a$ | $\Delta Ex^b$ | $PT_c^c$ | *P*-value | *q*-value[d] | Top 1%[e] | Top 5%[e] |
|---|---|---|---|---|---|---|---|---|
| ENSSSCG00000032094 | *DKK2* | -2.010 | -1.431 | -4.738 | 1.654E-05 | 3.830E-03 | x | x |
| ENSSSCG00000015334 | *PDK4* | -2.108 | -5.250 | -4.698 | 4.693E-03 | 1.330E-01 | x | x |
| ENSSSCG00000015037 | *IL18* | -1.655 | -1.191 | -3.682 | 4.787E-03 | 1.340E-01 | x | x |
| ENSSSCG00000005385 | *NR4A3* | -1.337 | -3.082 | -3.646 | 4.038E-02 | 4.098E-01 | x | x |
| ENSSSCG00000003766 | *DNAJB4* | -1.391 | -1.008 | -3.348 | 8.358E-03 | 1.905E-01 | x | x |
| ENSSSCG00000015969 | *CHRNA1* | -1.561 | -1.339 | -3.341 | 2.606E-03 | 9.406E-02 | x | x |
| ENSSSCG00000039419 | *SLCO4A1* | -1.055 | -2.279 | -3.180 | 2.820E-02 | 3.544E-01 | x | x |
| ENSSSCG00000049158 | | -1.107 | -1.096 | -3.164 | 3.182E-02 | 3.735E-01 | x | x |
| ENSSSCG00000004347 | *FBXL4* | -1.298 | -1.126 | -3.133 | 1.422E-03 | 6.520E-02 | x | x |
| ENSSSCG00000004979 | *MYO9A* | -1.239 | -1.003 | -3.043 | 7.296E-03 | 1.731E-01 | x | x |
| ENSSSCG00000013351 | *NAV2* | -1.163 | -1.196 | -2.863 | 2.605E-04 | 2.301E-02 | x | x |
| ENSSSCG00000032741 | *TBC1D9* | -0.913 | -1.061 | -2.736 | 1.534E-02 | 2.583E-01 | x | x |
| ENSSSCG00000031728 | *ABRA* | -1.238 | -1.393 | -2.704 | 1.295E-03 | 6.116E-02 | x | x |
| ENSSSCG00000006331 | *PBX1* | -0.891 | -1.039 | -2.480 | 1.135E-02 | 2.177E-01 | | x |
| ENSSSCG00000035037 | | -1.357 | -1.289 | -2.475 | 3.999E-03 | 1.212E-01 | | x |
| ENSSSCG00000038374 | *CIART* | -1.027 | -1.321 | -2.052 | 1.543E-02 | 2.587E-01 | | x |
| ENSSSCG00000023806 | *LRRN1* | -0.776 | -1.013 | -1.983 | 1.580E-01 | 7.074E-01 | | x |
| ENSSSCG00000009157 | *TET2* | -0.381 | -1.123 | -1.792 | 4.880E-01 | 9.582E-01 | | x |
| ENSSSCG00000011133 | *PFKFB3* | -0.022 | -2.256 | -1.785 | 9.712E-01 | 9.987E-01 | | x |
| ENSSSCG00000002283 | *FUT8* | -0.578 | -1.286 | -1.784 | 9.887E-02 | 6.059E-01 | | x |
| ENSSSCG00000023133 | *OSBPL6* | -0.432 | -1.088 | -1.772 | 3.835E-01 | 9.108E-01 | | x |
| ENSSSCG00000017986 | *NDEL1* | -0.767 | -1.644 | -1.759 | 1.006E-02 | 2.081E-01 | | x |
| ENSSSCG00000031321 | *NR4A1* | -0.630 | -1.328 | -1.720 | 6.298E-02 | 5.006E-01 | | x |
| ENSSSCG00000035101 | *KLF5* | -0.519 | -1.487 | -1.708 | 2.942E-01 | 8.488E-01 | | x |
| ENSSSCG00000004332 | *BACH2* | -0.714 | -2.105 | -1.705 | 9.089E-02 | 5.861E-01 | | x |
| ENSSSCG00000017983 | *PER1* | -0.773 | -1.073 | -1.627 | 3.000E-02 | 3.662E-01 | | x |

[a]$Log_2FC$: estimated $log_2$ fold change for mean exonic fractions from *gluteus medius* expression profiles of fasted *AL-T0* and fed *AL-T2* Duroc gilts; [b]ΔEx: exonic fraction increment ($Ex_2 – Ex_1$) when comparing exon abundances in *AL-T0* ($Ex_1$) vs *AL-T2* ($Ex_2$) Duroc gilts; [c]PTc: post-transcriptional signal (ΔEx – ΔInt) in z-score scale; [d]q-value: q-value calculated with the false discovery rate (FDR) approach [37]. [e]The "x" symbols represent the genes showing the top 1% (N = 13) and 5% (N = 26) negative $PT_c$ scores and with reduced ΔEx values > 2 folds (equivalent to ΔEx < -1 in the $log_2$ scale).

**Table 2:** microRNAs detected by edgeR as differentially expressed (|FC| > 1.5; *P*-value < 0.01) when comparing *lean* (N = 5) vs *obese* (N = 5) adipocyte expression profiles from Duroc-Göttingen minipigs according to their body mass index (BMI).

| miRNA | $log_2FC^a$ | *P*-value | *q*-value[b] | $log_2CPM^c$ obese | $log_2CPM^c$ lean |
|---|---|---|---|---|---|
| ssc-miR-92b-3p* | 1.854 | 3.647E-05 | 8.096E-03 | 7.263 | 9.004 |
| ssc-miR-190a | -2.617 | 3.285E-03 | 2.367E-01 | 3.810 | 0.889 |
| ssc-miR-148a-3p* | 0.825 | 3.834E-03 | 2.367E-01 | 11.096 | 11.875 |
| ssc-miR-204 | 1.797 | 4.395E-03 | 2.367E-01 | 4.923 | 6.866 |
| ssc-miR-1839-5p | -3.000 | 5.331E-03 | 2.367E-01 | 4.194 | 1.196 |
| ssc-miR-92a | 0.669 | 6.721E-03 | 2.487E-01 | 14.368 | 15.007 |
| ssc-miR-214-3p* | 0.874 | 9.042E-03 | 2.867E-01 | 9.116 | 9.977 |

[a]$Log_2FC$: estimated $log_2$ fold change for mean exonic fractions from adipocyte isolates belonging to lean and obese Duroc-Göttingen minipigs; [b]q-value: q-value calculated with the false discovery rate (FDR) approach [37]; [c]$Log_2CPM$: estimated $log_2$ counts-per-million (CPM) mean expression levels in *obese* and *lean* minipigs. *microRNAs verified by qPCR profiling.

**Table 3:** Genes with the top 5% post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 3 folds (equivalent to -1.58 in the $log_2$ scale) from adipocyte expression profiles of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI).

| mRNA | Gene | $log_2FC^a$ | $\Delta Ex^b$ | $PT_c^c$ | *P*-value | *q*-value[d] |
|---|---|---|---|---|---|---|
| ENSSSCG00000010814 | *ESRRG* | -0.591 | -5.305 | -6.425 | 7.364E-01 | 9.996E-01 |
| ENSSSCG00000032452 | *WFS1* | -2.198 | -2.138 | -5.510 | 9.509E-03 | 9.996E-01 |
| ENSSSCG00000039548 | *PTGFR* | -1.634 | -1.590 | -4.915 | 8.804E-03 | 9.996E-01 |
| ENSSSCG00000002265 | *FAM174B* | -1.244 | -1.726 | -4.179 | 5.385E-02 | 9.996E-01 |
| ENSSSCG00000016233 | *SERPINE2** | -1.735 | -2.060 | -3.603 | 5.684E-02 | 9.996E-01 |
| ENSSSCG00000006243 | *PENK* | -0.420 | -2.104 | -3.573 | 7.628E-01 | 9.996E-01 |
| ENSSSCG00000014921 | *PRSS23** | -1.141 | -1.739 | -3.360 | 2.719E-01 | 9.996E-01 |
| ENSSSCG00000017186 | *RNF157** | -1.218 | -2.338 | -3.317 | 2.413E-01 | 9.996E-01 |
| ENSSSCG00000031819 | *TP53I11* | -1.002 | -1.711 | -2.883 | 4.102E-01 | 9.996E-01 |
| ENSSSCG00000001089 | *GPLD1* | -0.872 | -1.761 | -2.723 | 4.302E-01 | 9.996E-01 |
| ENSSSCG00000003377 | *ACOT7* | -0.790 | -2.688 | -2.544 | 3.439E-01 | 9.996E-01 |
| ENSSSCG00000040464 | *LEP** | -0.747 | -2.186 | -2.463 | 1.880E-01 | 9.996E-01 |
| ENSSSCG00000025652 | *CDH1* | -0.472 | -2.592 | -2.372 | 6.533E-01 | 9.996E-01 |
| ENSSSCG00000011230 | *OSBPL10** | -0.576 | -1.594 | -1.869 | 4.272E-01 | 9.996E-01 |
| ENSSSCG00000017328 | *ARHGAP27* | -0.235 | -2.788 | -1.699 | 8.113E-01 | 9.996E-01 |

[a]$Log_2FC$: estimated $log_2$ fold change for mean exonic fractions from adipocyte expression profiles of *lean* and *obese* Duroc-Göttingen minipigs; [b]ΔEx: exonic fraction increment ($Ex_2$ – $Ex_1$) when comparing exon abundances in *obese* ($Ex_1$) vs *lean* ($Ex_2$) Duroc gilts; [c]PTc: post-transcriptional signal (ΔEx – ΔInt) in z-score scale; [d]q-value: q-value calculated with the false discovery rate (FDR) approach [37]. *mRNAs verified by qPCR profiling.

# Figure legends

**Figure 1:** Scatterplots depicting (**a**) transcriptional and post-transcriptional regulatory signals according to exonic (ΔEx) and intronic (ΔInt) fractions, (**b**), post-transcriptional regulatory signals according to exonic (ΔEx) and $PT_c$ (ΔEx – ΔInt) scores and (**c**), transient transcriptional regulatory signals according to intronic (ΔInt) and $PT_c$ (ΔEx – ΔInt) scores. (**d**) The classification and interpretation of transcriptional and post-transcriptional signals according to ΔEx, ΔInt ($T_c$) and ΔEx - ΔInt ($PT_c$) values. Enrichment analyses of the number of genes with the (**e**) top 1% and (**f**) top 5% negative post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds putatively targeted by upregulated miRNAs (FC > 1.5; *q*-value < 0.05) from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts and the consequences of incorporating context-based pruning of miRNA binding sites of type 8mer, 7mer-m8 and 7mer-A1. R: Raw enrichment analyses without any additional context-based pruning. AU: Enrichment analyses removing miRNA binding sites without AU-rich flanking sequences (30 nts upstream and downstream). M: Enrichment analyses removing miRNA binding sites located in the middle of the 3'-UTR sequence (45-55%). E: Enrichment analyses removing miRNA binding sites located too close (< 15 nts) to the beginning or the end of the 3'-UTR sequences. The black dashed line represents a *P*-value = 0.05. The red dashed line represents a *P*-value = 0.1.

**Figure 2:** Enrichment analyses of the number of binding sites (target-wise) of type 8mer, 7mer-m8 and 7mer-A1 for genes with the (**a**) top 1% and (**b**) top 5% negative post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds, putatively targeted by upregulated (FC > 1.5; *q*-value < 0.05), downregulated (FC < -1.5; q-value < 0.05) and top 5% most highly expressed miRNAs from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts. (**c**)

Enrichment analyses considering the significantly downregulated genes (FC < -2; q-value < 0.05) and excluding those with putative post-transcriptional regulatory signals. The black dashed line represents a *P*-value = 0.05. The red dashed line represents a *P*-value = 0.1.

**Figure 3:** (**a**) Scatterplot depicting the top 1% negative post-transcriptional regulatory signals according to exonic (ΔEx) and $PT_c$ (ΔEx – ΔInt) scores and highlighted genes (in purple) putatively targeted by upregulated (FC > 1.5; *q*-value < 0.05) miRNAs from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts. (**b**) Enrichment analyses of the number of genes with the top 1% negative post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds putatively targeted by upregulated miRNAs (FC > 1.5; *q*-value < 0.05), downregulated miRNAs (FC < -1.5; q-value < 0.05) and the top 5% most highly expressed miRNAs. (**c**) Scatterplot depicting the top 5% negative post-transcriptional regulatory signals according to exonic (ΔEx) and $PT_c$ (ΔEx – ΔInt) scores and highlighted genes (in purple) putatively targeted by upregulated (FC > 1.5; *q*-value < 0.05) miRNAs. (**d**) Enrichment analyses of the number of genes with the top 5% negative post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds putatively targeted by miRNAs. (**e**) Scatterplot depicting differentially upregulated (in green) and downregulated (in red) genes (|FC| > 2; *q*-value < 0.05) according to exonic (ΔEx) and $PT_c$ (ΔEx – ΔInt) scores. (**f**) Enrichment analyses considering the significantly downregulated genes (FC < -2; *q*-value < 0.05) and excluding those with putative post-transcriptional regulatory signals. The black dashed line represents a *P*-value = 0.05. The red dashed line represents a *P*-value = 0.1.

**Figure 4:** Covariation enrichment scores (CES) for the exonic and intronic fractions of genes with the (**a**) top 1% (N =12) and (**b**) top 5% (N = 18) negative post-transcriptional

signals ($PT_c$) and reduced exonic fractions (ΔEx) > 2 folds, putatively targeted by upregulated miRNAs (FC > 1.5; *q*-value < 0.05). The control set of genes were generated by permuted (N = 1,000) sets of genes of length N = 12 for CES in (**a**) and N = 18 for CES in (**b**).

**Figure 5:** (**a**) Scatterplot depicting the top 1% transcriptional regulatory signals according to intronic (ΔInt) and $PT_c$ (ΔEx – ΔInt) scores and highlighted genes (in blue) with putative transient transcription upregulation from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts. (**b**) Depletion analyses of the number of genes with the top 1% negative post-transcriptional signals ($PT_c$) and increased intronic fractions (ΔInt) > 2 folds putatively targeted by upregulated miRNAs (FC > 1.5; *q*-value < 0.05), downregulated miRNAs (FC < -1.5; q-value < 0.05) and the top 5% most highly expressed miRNAs. (**c**) Scatterplot depicting the top 5% negative post-transcriptional regulatory signals according to intronic (ΔInt) and $PT_c$ (ΔEx – ΔInt) scores and highlighted genes (in blue) with putative transient transcription upregulation. (**d**) Depletion analyses of the number of genes with the top 5% negative post-transcriptional signals ($PT_c$) and increased intronic fractions (ΔInt) > 2 folds putatively targeted by miRNAs. (**e**) Scatterplot depicting differentially upregulated (in green) and downregulated (in red) genes (|FC| > 2; *q*-value < 0.05) according to intronic (ΔInt) and $PT_c$ (ΔEx – ΔInt) scores. (**f**) Depletion analyses considering the significantly upregulated genes (FC > 2; *q*-value < 0.05) and excluding those with putative transient transcriptional upregulation signals. The black dashed line represents a *P*-value = 0.05. The red dashed line represents a *P*-value = 0.1.

**Figure 6:** (**a**) Scatterplot depicting the top 5% negative post-transcriptional regulatory signals according to exonic (ΔEx) and $PT_c$ (ΔEx – ΔInt) scores and highlighted genes (in purple) putatively targeted by upregulated (FC > 1.5; *P*-value < 0.01) miRNAs from

adipocyte expression profiles of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI). (**b**) Enrichment analyses of the number of genes with the top 5% negative post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 3 folds (equivalent to 1.58 in the $log_2$ scale) putatively targeted by upregulated miRNAs (FC > 1.5; *P*-value < 0.01), downregulated miRNAs (FC < -1.5; *P*-value < 0.01) and the top 5% most highly expressed miRNAs. (**c**) Barplots depicting qPCR $log_2$ transformed relative quantities (Rq) for *LEP*, *OSBPL10*, *PRSS23*, *RNF157* and *SERPINE2* mRNA transcripts measured in adipocytes from the retroperitoneal fat of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs. (**d**) Barplots depicting qPCR $log_2$ transformed relative quantities (Rq) for ssc-miR-148a-3p, ssc-miR-214-3p and ssc-miR-92b-3p miRNA transcripts measured in the isolated adipocytes from the retroperitoneal fat of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs.

# Supplementary Tables

**Table S1:** Phenotype values for selected Duroc-Göttingen minipigs from the F2-UNIK source population according to their body mass indexes (BMI).

**Table S2:** Primers for qPCR validation of selected mRNA and miRNA genes according to EISA results in the F2-UNIK Duroc-Göttingen minipig population comparing *lean* (N = 5) and *obese* (N = 5) individuals.

**Table S3:** Raw Cq values after efficiency correction measuring adipocyte expression profiles of selected mRNAs and miRNAs from *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs.

**Table S4:** mRNA genes detected by *edgeR* as differentially expressed when comparing *gluteus medius* expression profiles of fasted *AL-T0* (N = 11) and fed *AL-T2* (N = 12) Duroc gilts.

**Table S5:** microRNA genes detected by *edgeR* as differentially expressed when comparing *gluteus medius* expression profiles of fasted *AL-T0* (N = 11) and fed *AL-T2* (N = 12) Duroc gilts.

**Table S6:** EISA analyses for transcriptional and post-transcriptional signals detected in *gluteus medius* skeletal muscle expression profiles of fasted (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

**Table S7:** Genes with the top 1% and 5% post-transcriptional signals ($PT_c$) and increased intronic fractions (ΔInt) > 2 folds (equivalent to 1 in the $log_2$ scale) from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

**Table S8:** Binding sites for differentially upregulated miRNAs from mRNA genes with the top 1% and 5% negative $PT_c$ scores and reduced ΔEx > 2 folds from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

**Table S9:** Binding sites for differentially upregulated miRNAs from the 5'-UTR of the two transcripts annotated for the ENSSSCG00000049158 mRNA gene.

**Table S10:** Binding sites for differentially upregulated miRNAs from mRNA genes with the top 1% and 5% negative $PT_c$ scores and increased ΔInt > 2 folds from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

**Table S11:** Transcription factor binding sites within promoter region of mRNA genes with the top 1% negative $PT_c$ scores and increased ΔInt > 2 folds from *gluteus medius*

skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts.

**Table S12:** mRNA genes detected by *edgeR* as differentially expressed when comparing adipocyte expression profiles from *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI).

**Table S13:** microRNA genes detected by *edgeR* as differentially expressed when comparing adipocyte expression profiles from *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI).

**Table S14:** Genes with the top 5% post-transcriptional signals ($PT_c$) and reduced exonic fractions (ΔEx) > 3 folds (equivalent to -1.58 in the $log_2$ scale) from adipocyte expression profiles of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI).

**Table S15:** Binding sites for differentially upregulated miRNAs from mRNA genes with top 5% negative $PT_c$ scores and reduced ΔEx > 3 folds from adipocyte expression profiles of *lean* (N = 5) and *obese* (N = 5) Duroc-Göttingen minipigs according to their body mass index (BMI).

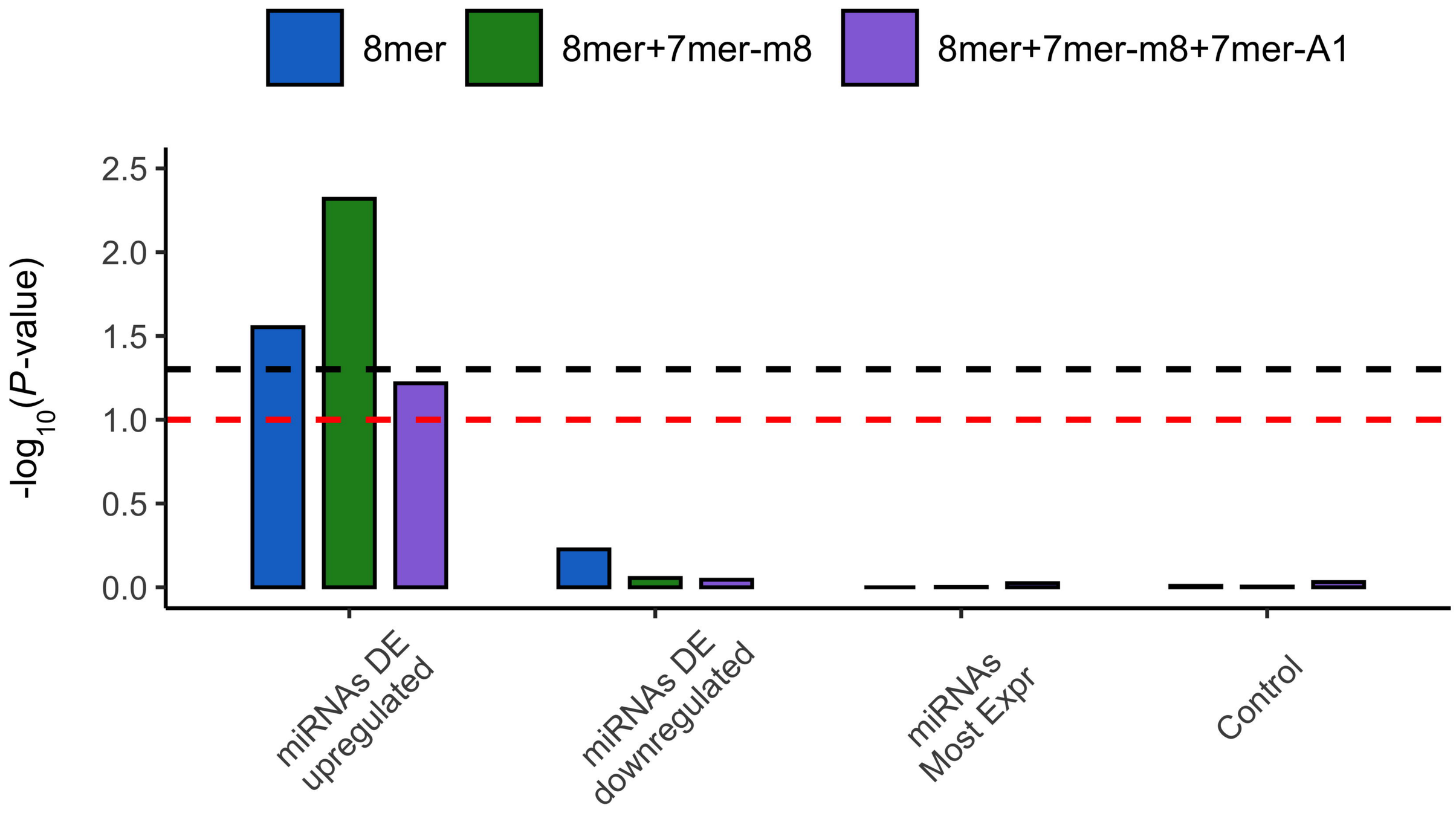
## Supplementary Figures

**Figure S1:** Scatterplots depicting the exonic (ΔEx) and intronic (ΔInt) fractions from *gluteus medius* skeletal muscle expression profiles of fasting (*AL-T0*, N = 11) and fed (*AL-T2*, N = 12) Duroc gilts. (**a**) mRNA genes with top 5% post-transcriptional ($PT_c$) negative scores and reduced exonic (ΔEx) fractions > 2 folds (equivalent to -1 in the $log_2$ scale), suggestive of miRNA-driven post-transcriptional regulation. (**b**) mRNA genes with top 5% post-transcriptional ($PT_c$) scores and increased intronic (ΔInt) fractions > 2

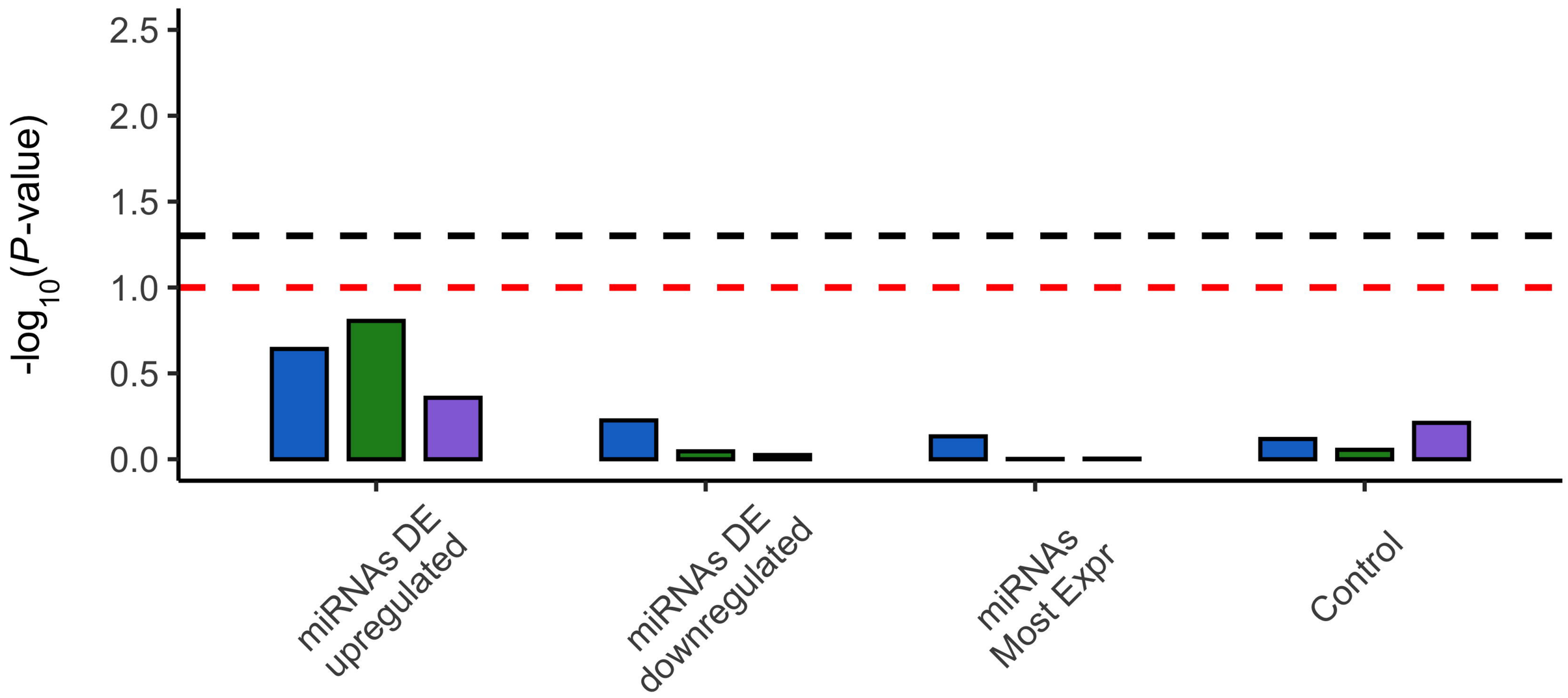
folds (equivalent to 1 in the $\log_2$ scale), suggestive of transient TF-driven transcriptional regulation. (**c**) mRNA genes differentially expressed showing upregulation (in green) and downregulation (in red) in fed (*AL-T2*, N = 12) Duroc gilts with respect to their fasted (*AL-T0*, N = 11) counterparts.

**Figure S2:** Transcription factor (TF) binding motif enrichment analyses for (a) mRNA genes with top 5% post-transcriptional ($PT_c$) negative scores and increased intronic (ΔInt) fractions > 2 folds (equivalent to -1 in the $\log_2$ scale), suggestive of transient TF-driven transcriptional regulation. (**b**) mRNA genes with top 5% transcriptional ($T_c$) scores and increased exonic (ΔEx) fractions > 2 folds (equivalent to 1 in the $\log_2$ scale), suggestive of mRNA upregulation. (**c**) mRNA genes differentially expressed showing upregulation (FC > 2; *q*-value < 0.05) in fed (*AL-T2*, N = 12) Duroc gilts with respect to their fasted (*AL-T0*, N = 11) counterparts.

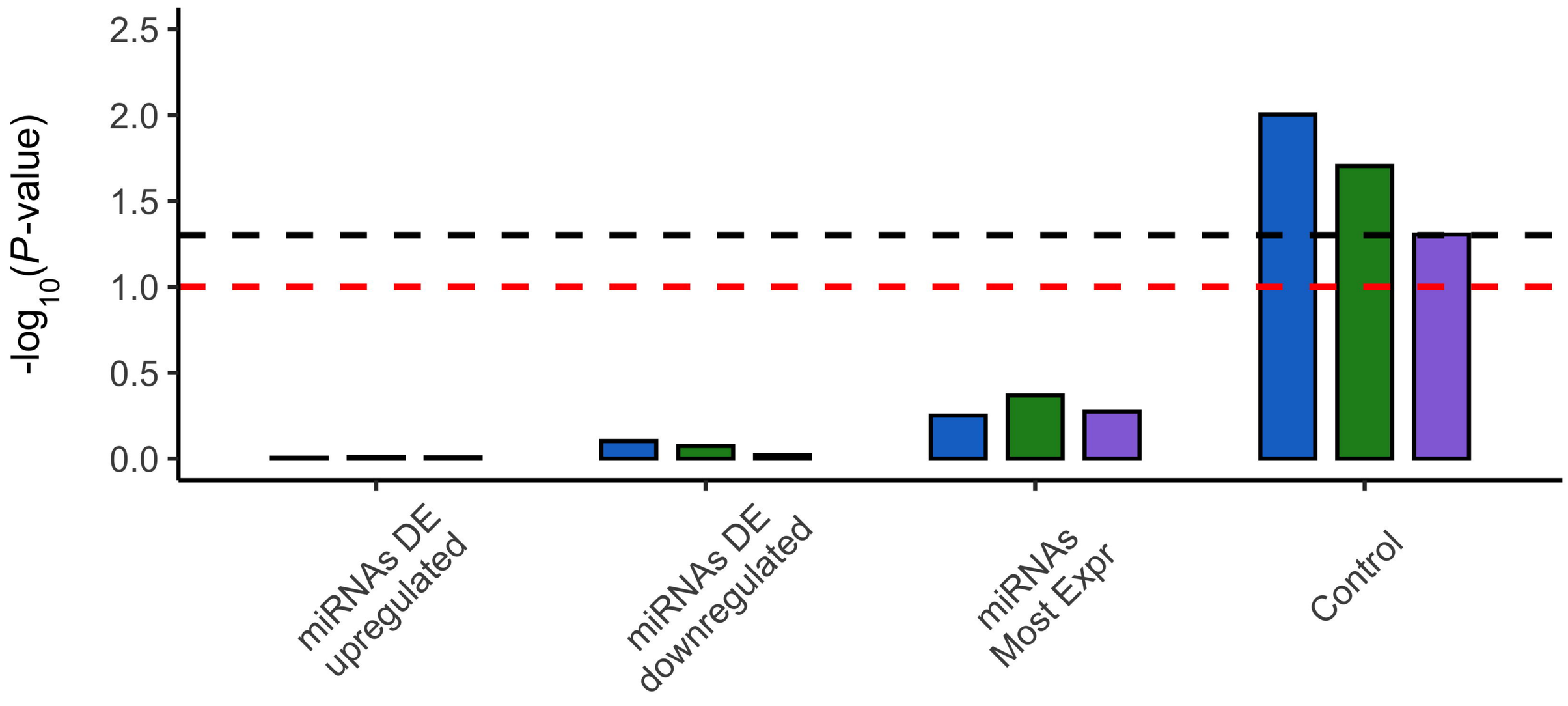
a

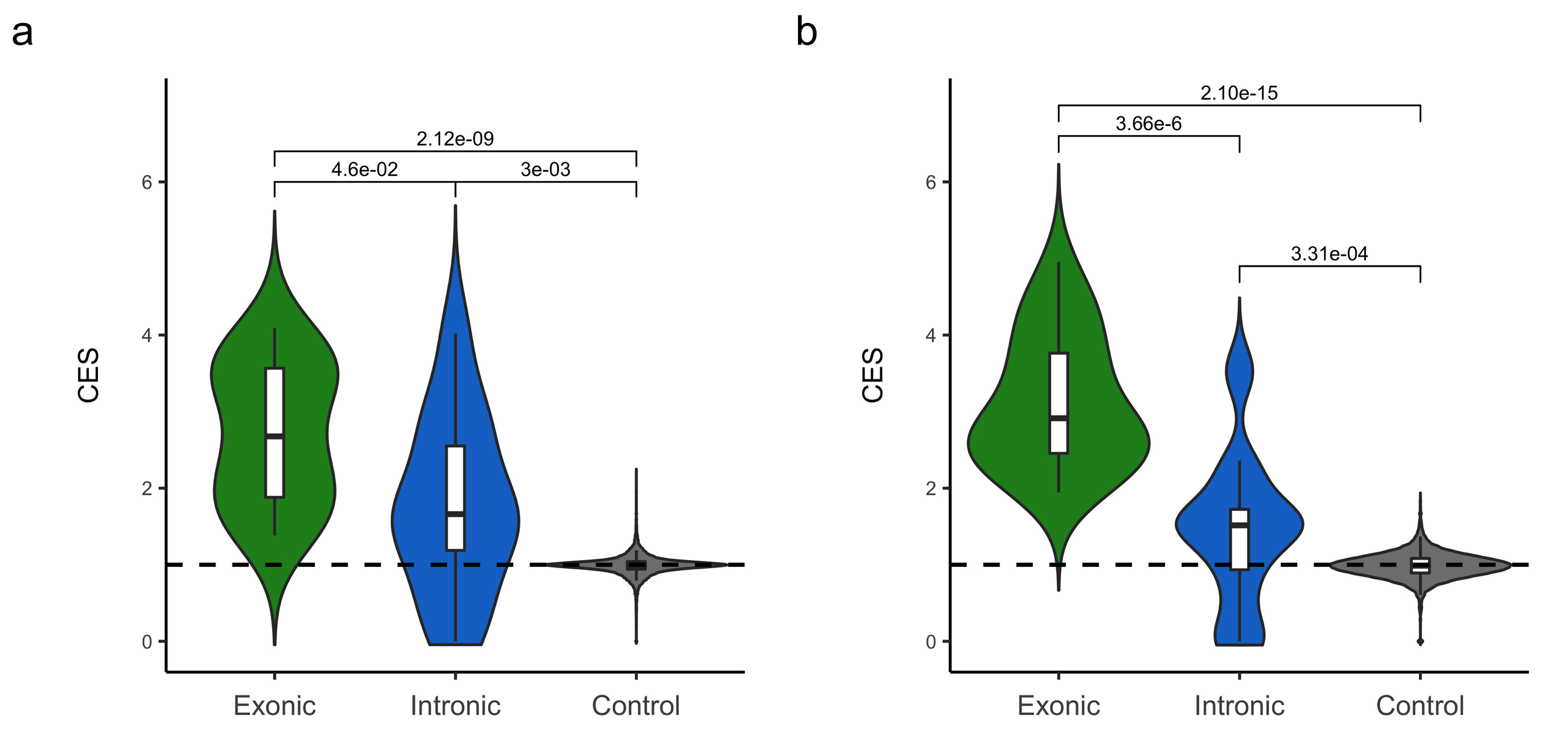

8mer
8mer+7mer-m8
8mer+7mer-m8+7mer-A1
-log10(P-value)
2.5
2.0
1.5
1.0
0.5
0.0
miRNAs DE upregulated
miRNAs DE downregulated
miRNAs Most Expr
Control

b


-log10(P-value)
2.5
2.0
1.5
1.0
0.5
0.0
miRNAs DE upregulated
miRNAs DE downregulated
miRNAs Most Expr
Control

c


-log10(P-value)
2.5
2.0
1.5
1.0
0.5
0.0
miRNAs DE upregulated
miRNAs DE downregulated
miRNAs Most Expr
Control


a
2.12e-09
4.6e-02
3e-03
CES
0
2
4
6
Exonic
Intronic
Control
b
2.10e-15
3.66e-6
3.31e-04
CES
0
2
4
6
Exonic
Intronic
Control