Integrated analysis of proteomic and transcriptomic data highlights late fetal muscle maturation process

Valentin Voillet, Magali San Cristobal, Marie-Christine Pere, Yvon Billon, Laurianne L. Canario, Laurence Liaubet, Louis Lefaucheur

MCP Papers in Press. Published on January 8, 2018 as Manuscript M116.066357

Integrated Analysis of Proteomic and Transcriptomic Data Highlights Late Fetal Muscle Maturation Process

Valentin Voillet[1], Magali San Cristobal[1], Marie-Christine Père[2], Yvon Billon[3], Laurianne Canario[1], Laurence Liaubet[1] and Louis Lefaucheur[2,*]

[1] **Université de Toulouse, INRA, INPT, INP-ENVT, UMR1388, GenPhySE, F-31326 Castanet-Tolosan, France**

[2] **INRA, UMR1348, PEGASE, F-35590 Saint-Gilles, France**

[3] **INRA, UE1372, GenESI, F-17700 Surgères, France**

∗ **E-mail: louis.lefaucheur@inra.fr**

Running title: Integration of omics data defines muscle maturity

Abbreviations

ACADVL: Acyl-CoA dehydrogenase very long chain, mitochondrial

AMPK: AMP-activated protein kinase

ANXA2: Annexin A2

ATP5A1: ATP synthase subunit alpha, mitochondrial

BP: Biological process

CC: Cellular component

CKM: Creatine kinase, muscle, cytoplasm

CKMT2: Creatine kinase, mitochondrial 2

CS: Citrate synthase

DDAH1: Dimethylarginine dimethylaminohydrolase

ESR1: Estrogen receptor alpha

FDR: False discovery rate

FETUB: Fetuin-B, cystein protease inhibitor family

GO: Gene ontology

GPD1: Glycerol-3-phosphate dehydrogenase 1, cytoplasmic

GSN: Gelsolin

HAD: Beta-hydroxy-Acyl-CoA dehydrogenase

HPRT1: Hypoxanthine phosphoribosyltransferase 1

KCNJ11: Potassium Channel, Inwardly Rectifying Subfamily J, Member 11

LDB3: LIM domain binding 3

LDH: Lactate dehydrogenase

LM: Longissimus muscle

LW: Large White

MACROD1: MACRO domain-containing protein 1

MF: Molecular function

MS: Meishan

MyHC: Myosin heavy chain

OXCT1: Succinyl-CoA:3-ketoacid-coenzyme A transferase 1, mitochondrial

PCIT: Partial correlation information theory

PCK2: Phosphoenolpyruvate carboxykinase 2, mitochondrial

PDIA3: Protein disulfide-isomerase A3

PPARGC1A: Peroxisome proliferator-activated receptor gamma coactivator 1-alpha

PSMC5: Proteasome 26S regulatory subunit

PYGM: Muscle glycogen phosphorylase

RF: Random forest

sCCA: Sparse canonical correspondance analysis

SEPT2: Septin 2

SIRT1: Sirtuin 1

sPLS: Sparse partial least square

sPLS-DA: Sparse partial least square - discriminant analysis

TNNT3: Troponin T type 3

Summary

Background. In pigs, the perinatal period is the most critical time for survival. Piglet maturation, which occurs at the end of gestation, leads to a state of full development after birth. Maturity is thus an important determinant of early survival. Skeletal muscle plays a key role in adaptation to extra-uterine life, e.g. motor function and thermoregulation. Progeny from two breeds with extreme neonatal mortality rates were analyzed at 90 and 110 days of gestation (dg). The Large White breed is a highly selected breed for lean growth with a high rate of mortality at birth, whereas the Chinese Meishan breed is fatter and more robust and has a low mortality rate. The aim of our study was to identify important molecular signatures underlying late fetal muscle development of longissimus muscle.

Method. First, integrated analysis was used to explore relationships between co-expression network models built from a proteomic dataset (bi-dimensional electrophoresis), and biological phenotypes of interest. Second, possible correlations with a transcriptomic dataset (microarrays) were investigated to combine different layers of expression with a focus on transcriptional regulation.

Results. Muscle glycogen content and myosin heavy chain polymorphisms were found to be good descriptors of muscle maturity and were used for further data integration analysis. Using 113 protein spots (89 identified unique proteins), network inference, correlation with biological phenotypes and functional enrichment revealed that mitochondrial oxidative metabolism was a key determinant of neonatal muscle maturity. Some proteins, including ATP5A1 and CKMT2, were identified as important nodes in the network related to muscle metabolism. Transcriptomic data suggest that over-expression of mitochondrial PCK2 was involved in the greater glycogen content of Meishan fetuses at 110 dg. GPD1, an enzyme involved in the mitochondrial oxidation of cytosolic NADH, was also over-expressed in the Meishan breed. Thirty-one proteins exhibited a positive correlation between their mRNA and protein levels, suggesting transcriptional regulation in both extreme fetal genotypes. Gene ontology enrichment analysis and Ingenuity analysis identified *PPARGC1A* and *ESR1* as possible transcriptional factors positively involved in late fetal metabolic maturation of skeletal muscle.

Introduction

One objective of systems biology is to investigate the regulation and interaction of various components of the cell including DNA (genomics) [1], mRNA (transcriptomics) [2], proteins (proteomics) [3] or metabolites (metabolomics) [4]. Even though transcriptomic analysis provides deep insights into cellular processes, possible conclusions are limited [5]. Indeed, mRNA expression is not always a good predictor of protein level because low correlations between mRNA and protein expression levels are often observed. For example, a relatively small change in mRNA expression can result in a major change in the total abundance of the corresponding protein, enabling potentially different conclusions to be drawn from transcriptomic and proteomic analyses.

Biological integration of transcriptome and proteome remains challenging in omics studies due to the marked differences between the two approaches, e.g. the dynamic range of regulation, incomplete annotation or isoform differences [6]. The expression of genes may not be correlated with the abundance of proteins and provide no information on post-transcriptional events, e.g. translational efficiency, alternative splicing, folding or assembly into complexes [5, 7, 8]. An integrated multi-omics approach, with gene expression experiments and large-scale protein identification experiments, should provide a deeper understanding of the functional interactions between mRNA and protein layers as a complex biological system. Several biological data integration strategies have already been suggested, see review by Ritchie et al. (2015) [9]. In this review, the authors suggested three meta-dimensional analyses combining multiple data types in the same analysis: concatenation, transformation and model-based integration. Here we chose to develop an innovative integration strategy combining state-of-the-art statistical and computational methods using proteomic and phenotypic data, with the incorporation of transcriptomic information, to observe and identify some important proteins with possible transcriptional regulation. One of the advantages of our strategy is to avoid the scale issues that may arise when different types of data are combined.

Networks are increasingly used as tools for analysis and for the visualization of data in biology and genetics [10–13]. A complex network architecture into clusters of functionally related genes/proteins can be explored and genes/proteins with high connectivity (called hubs) can be identified. In our study, integrated network analysis was first performed to explore relationships between co-expression network models,

built from a proteomic dataset, and phenotypes of interest that would enable identification of important molecular signatures underlying late fetal muscle development. Correlations with a transcriptomic dataset were then investigated to complete and combine different layers of expression.

Here we report the results of multi-omics analyses of muscle during the last three weeks of gestation in pigs. The objective was to identify proteins, with possible transcriptional regulation, and related biological mechanisms, specially those involved in differences in the muscle maturation process in late gestation between two extreme pig breeds: Large White and Meishan, and reciprocal crosses. The Large White (LW) breed is a highly selected breed for lean growth with a high rate of mortality at birth, whereas the Chinese Meishan (MS) is a fatter and more robust breed that produces piglets with an extremely low mortality rate [14, 15]. Physiological muscle maturity, which occurs at the end of gestation, has already been shown to improve early survival after birth [16–18]. Successful maturation in late gestation thus likely leads to a state of full development and promotes early survival after birth. Postnatal mortality due to immaturity is not only an issue in pigs but affects other mammals including sheep [19] and humans [20, 21]. Adaptation to extra-uterine life is therefore a major factor in the survival of all mammal species. In the present study, using network data integration analysis, we identified key proteins involved in piglet maturity during late gestation and provided an overview of the muscle maturation process in late gestation of which many aspects can be generalized to other mammals.

Materials and Methods

Source data

Ethics statement

Use of animals and procedures performed in this study was approved by the European Union legislation (directive 86/609/EEC) and French legislation in the Midi-Pyrénées Region of France (Decree 2001-464 29/05/01; accreditation for animal housing C-35-275-32). The technical and scientific staff obtained individual accreditation (MP/01/01/01/11) from the ethics committee (region Midi-Pyrénées, France) to experiment on living animals. All the fetuses used in this study were males and were obtained by caesarean.

Study design

Details regarding animal resources and experimental designs can be found in [18]. Briefly, *longissimus skeletal muscle* mRNA, protein and phenotypic data were acquired at two developmental time points (90 and 110 days of gestation (dg)) from four fetal genotypes. These genotypes consisted in two extreme breeds concerning mortality at birth (Meishan (MS) and Large White (LW)) and two crosses (MSLW from LW sows and LWMS from MS sows). MS and LW sows were inseminated with mixed semen so that each litter was composed of purebred and crossbred fetuses. The two developmental time points correspond to the end of gestation when intense maturation of muscle fibers occurs from 90 dg to birth (around 114 dg) [16, 17, 22]. Therefore, eight conditions were considered according to the two development time points and the four fetal genotypes (n = 64 fetuses exhibiting birth weight close to the average birth weight within litter and genotype, $n_{d90.MS} = 8$, $n_{d90.LW} = 8$, $n_{d90.MSLW} = 8$, $n_{d90.LWMS} = 8$, $n_{d110.MS} = 8$, $n_{d110.LW} = 10$, $n_{d110.MSLW} = 6$ and $n_{d110.LWMS} = 8$).

Muscle sampling and biochemical analysis

The *longissimus muscle* (LM) was collected at the last rib level within 30 min after death, cut into small pieces, snap frozen in liquid nitrogen, and stored at -75°C until further analyses. Glycogen content was determined in LM according to the method described by Good et al. [23] with minor adaptations [24], and is expressed in g/100 g tissue wet weight. Myosin heavy chain (MyHC) polymorphism was characterized both at the mRNA level by quantitative real-time PCR amplification using the TaqMan technology and

at the protein level using one dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE) as previously described in [25]. At the mRNA level, the expression of each MyHC (*MYH7* = type I MyHC, *MYH2* = type IIa MyHC, *MYH1* = type IIx MyHC, *MYH4* = type IIb MyHC, *MYH3* = embryonic MyHC, *MYH8* = perinatal MyHC, *MYH6* = α-cardiac MyHC and *MYH13* = extraocular MyHC) was calculated based on the PCR efficiencies and a calibrator, and expressed in comparison to an invariant endogenous reference gene (hypoxanthine phosphoribosyltransferase 1, *HPRT1*) as described by Pfaffl [26]. *HPRT1* expression was not affected by the fixed factors used in the statistical analysis. Types I, IIa, IIx and IIb MyHC are the four MyHC that define muscle fiber types in adult skeletal muscle [27], whereas embryonic, perinatal, and α-cardiac MyHC are expressed transitorily during the fetal and early postnatal periods in pigs [28, 29]. At the protein level, four bands were separated by 1D SDS-PAGE and corresponded to the embryonic, perinatal, fast-type II (IIa + IIx + IIb) and slow-type (I + α-cardiac) MyHC, respectively [30]. Each band is expressed as a percentage of all four bands within a lane.

Proteome analysis

Muscle total protein extraction and bi-dimensional electrophoresis were performed on the 64 muscle samples as previously described in [31]. Briefly, total proteins were extracted from 50 mg of powdered frozen muscle tissue homogenized into 1 mL lysis buffer containing 7.5 M urea, 2 M thiourea, 2% CHAPS, 40 mM Tris base, 50 mM DTT, 1 mM EDTA, 1% IPG buffer pH 3-11 non-linear (NL), 1 μg/mL pepstatin and 1.5 mg/mL complete protease inhibitor cocktail (Roche, Mannheim, Germany), using a handheld Teflon-pestle-glass Potter-Elvehjem homogenizer. The lysate was incubated at room temperature for 1 h with rotational shaking, followed by centrifugation at 60,000g for 1 h at 18°C. Supernatant was collected avoiding the fat layer, aliquoted and stored at -75°C until further analysis. Protein concentration in the supernatants was determined using the RC DC Protein Assay (Bio-Rad, Hercules, CA, USA) with bovine serum albumin as a standard. For the first dimension, 300 μg protein were loaded onto immobilized pH gradient strips (pH 3-11 NL, GE Healthcare, Uppsala, Sweden) and isoelectric focusing (IEF) was performed using an Ettan IPGphorII system (GE Healthcare) at 20°C up to a total of 88,600 Vh. At the completion of IEF, equilibrated strips were transferred onto the top of a 12.5% uniform SDS-PAGE gel using a vertical Ettan DALTsix system (GE Healthcare). After migration, gels were stained with Coomassie Brilliant Blue G-250 (Bio-Rad). The gels were scanned using an UMAX ImageScanner (GE Healthcare) and spot detection and quantification were performed by image analysis (Melanie 2D gene

analysis software V7.0; Swiss Institute of Bioinformatics, Lausanne, Switzerland). Artefacts and saturated spots were removed from the analysis. Spots were matched across all 64 samples and 1,025 valid spots were successfully matched between gels. For each gel, each spot volume was expressed as a percentage of the volume of all matched spots on a given gel. Data were further log2 transformed. Representative 2D electrophoresis gels are included in Additional file 1 and Additional file 2.

Experimental Design and Statistical Rationale

Protein identification is a really *a posteriori* time-consuming work. Statistics were done in a complete blind manner and about 200 spots were expected for mass spectrometry identification. So that different statistical methods were chosen to obtain a wide spectrum of protein expression including differential analyses for depicting biological processes and discriminant analyses to identify possible markers. The purpose of those methods was not the same. The analyses of variance were performed to get differentially expressed spots according to the gestational time points and the fetal genotypes. Random forests and sPLS-DA were used to find supplementary spots with a predictive power for gestational stages and/or fetal genotypes. Finally, several statistical multivariate methods (sPLS and sCCA) were also performed to find additional spots correlated with phenotypic biological characters of interest such as MyHC profiles. These methods included analyses of variance to detect spots significantly affected by gestational time points and fetal genotypes, in an additive or non-additive manner, as described in [18]. Briefly, to analyze jointly differences between breeds and gestational ages, the following mixed linear model was first fitted to each spot: $Y_{ijk} = \mu + A_i + FG_j + A.FG_{ij} + S_k + \epsilon_{ijk}$, with $i \in \{d90, d100\}$, $j \in \{LW, MS, LWMS, MSLW\}$, $k = 1, ...18$, $S_k \sim \mathcal{N}(\mu, \sigma_e^2)$ independent and identically distributed (iid) and $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$ iid. S_k and ϵ_{ijk} are mutually independent. Y_{ijk} is the expression of the spot being studied, μ a general mean of the considered gene expression and ϵ_{ijk} is a residual. This model includes two fixed effects and their interaction: A_i is the effect of fetal gestational age i, FG_j the effect of fetal genotype j and $A.FG_{ij}$ the interaction effect between gestational age i and genotype j. S_k represents the random sow effect. A F-type test was achieved by comparing the complete model and the reduced model: $Y_{ijk} = \mu + S_k + \epsilon_{ijk}$. A correction for multiple testing was then implemented using False Discovery Rate (FDR). The list of Differentially Expressed Spots (DESs) was thereafter partitioned into 4 sub-models. Sub-model 1 combined the two fixed effects and their interaction. Sub-model 2 involved the two fixed effects in an additive manner. Sub-model 3 included only the fetal gestational age effect, whereas sub-model 4 contained only

the fetal genotype effect. All models included the random sow effect. The Bayesian Information Criterion (BIC) was used to associate each DES with one of these four sub-models. Only spots from sub-models 1 and 2 were kept for identification.

Secondly, random forest (RF) [32] and sparse partial least square discriminant analysis (sPLS-DA) [33]) were performed to find supplementary spots with a predictive power for gestational stages and/or fetal genotypes. Several RFs analyses were conducted. We computed a classification RF according to the experimental design and we also performed some regression RFs using phenotypes of interest (such as embryonic or adult fast myosins). In each case, we selected spots with a high stability according to several importance criteria (Gini and Accuracy indexes for the classification RF, Accuracy and the mean decrease in MSE for regression RFs). We performed 20 RFs and kept the first fifteen spots according to those importance criteria in order to evaluate the stability of the variable selection procedure. This allowed us to assess how often a given spot, selected with our parameters/criteria, is selected when running several RF procedures. In sPLS-DA, we chose to keep two axes and to keep 25 spots by axis. The first axis discriminated the fetuses according to the gestational age, whereas the second one discriminated fetuses according to the fetal genotype. No parameter was tuned using the *mixOmics* R functions because we wanted to keep about 50 spots from this specific analysis and only two axes based on the PCA obtained using the normalized proteome dataset.

Finally, several statistical multivariate methods (sparse partial least square (sPLS) [34] and sparse canonical correlation analysis (sCCA) [35]) were also performed to find additional spots correlated with phenotypic biological characters of interest such as MyHC profiles, plasmatic values and muscle glycogen content. In sPLS analysis, 20 and 30 spots were kept according to axis 1 and 2, respectively. In sCCA, about 50 spots were also kept to get as many spots as with sPLS analysis. Like in sPLS-DA, the first axis discriminated the fetuses according to the gestational age and the second one according to the fetal genotype. In total, 179 spots were selected and manually excised from preparative gels loaded with 600 μg of proteins from pooled samples for further identification by nano-LC-MS/MS, as described in [36].

An R script is provided as additional file in order to allow outside labs to reproduce all these analyses (Additional File 3). Raw data that is used in the R script can be found in supplementary files *gel2D_phenoData_Voillet_2016.csv* and *gel2D_Voillet_2016.csv*.

Sequencing grade modified trypsin (Promega, ref. V511A) was used to generate peptides. In-gel tryptic

digestion and protein identification by mass spectrometry were performed at the proteomic facilities in Clermont-Ferrand (PFEMcp, INRA, Clermont-Ferrand Theix, France). Tryptic peptides were analyzed by nano LC-MS/MS using nano-LC system Ultimate 3000^{TM} RSLC (Dionex, Voisins le Bretonneux, France) coupled on-line to an LTQ VELOS mass spectrometer (ThermoFisher Scientific, Courtaboeuf, France) operated in a CID top 5 mode (*i.e.* one full scan MS and the five major peaks in the full scan were selected for MS/MS). For database searches and protein identification, Thermo Proteome Discoverer 1.4 software was used with Mascot (Mascot server v2.2, http://www.matrixscience.com) to submit MS/MS data to the SwissProt sequence database restricted to *Sus scrofa* (UniP Sus scrofa, 26127 sequences, June 6th, 2014 release). The following parameters were chosen for the searches: the mass tolerance for parent and fragment ions was set to 1.5 Da and 0.5 Da, respectively, and a maximum of two missed cleavages was allowed. Variable modifications were methionine oxidation (M) and carbamidomethylation (C) of cysteine. Results were scored using the probability-based Mowse algorithm, where the protein score is -10*log(P) and P is the probability that the observed match is a random event. Protein identifications were considered valid if at least two peptides with a statistically significant Mascot score > 36 were assigned, and at least 20% sequence coverage was required. The accuracy of the experimental to theoretical isoelectric point and molecular weight were also considered. Among the 179 selected spots, 120 spots, corresponding to 89 unique proteins, were successfully identified (Additional Files 4 and 5). Raw mass spectrometry data have been deposited in the publicly accessible repository MassIVE site (ftp://MSV000081117@massive.ucsd.edu/ with id0181 as the password).

SDS-PAGE and immunoblotting

An aliquot of extracted proteins as used in the proteomic analysis was denatured in 1X Laemmli solution (Sigma-Aldrich, St. Louis, MO, USA, ref S3401) at 95°C during 5 min. Equal amounts of protein (30 μg per lane) were loaded onto 1.5 mm thick polyacrylamide gels (4 and 10% acrylamide in the stacking and resolving gels, respectively) set on a Bio-Rad Mini-Protean tetra electrophoretic system as previously described [37]. After migration in 1X running buffer (Bio-Rad, Marnes-la-Coquette, France, ref. 161-0772), separated proteins were transferred overnight at 4°C to polyvinylidene fluoride membranes at 30 V using 1X transfer buffer (Bio-Rad, ref. 161-0771). Membranes were blocked for an hour in 1X TBS-T with 5% blocking agent (Bio-Rad, ref 170-6404) at room temperature, and immunoblotted overnight at 4°C with primary monoclonal antibody specific for CKMT2 (Abcam, Cambridge, MA, USA, ab131179,

1:1,000 dilution). Blots were washed five times for 5 min with TBS-T before incubation for 1 h at room temperature with the anti-mouse (from sheep) immunoglobulin G horseradish peroxidase conjugated secondary antibody (GE Healthcare, Uppsala, Sweden, 1:5,000 dilution). Finally, the membranes were washed with TBS-T, followed by development using an Enhanced Chemiluminescence Western blotting kit (ECL-Prime, GE Healthcare). Images were scanned with an ImageQuant LAS 4000 (GE Healthcare) and protein bands were quantified using the ImageQuant TL program. Membrane normalization was based on a common sample repeated on every gel in three replicates and CKMT2 expression was expressed as arbitrary units (AU) based on the ratio between CKMT2 signal of a given experimental sample and CKMT2 signal of the common sample within the membrane. Beta-actin and tubulin could not be used as normalizing factor because their expression was highly variable and influenced by the experimental factors.

Metabolic enzyme activities

Activities of lactate dehydrogenase (LDH, E.C. 1.1.1.2.7), citrate synthase (CS, E.C. 1.1.3.7) and beta-hydroxy-acyl-CoA dehydrogenase (HAD, E.C. 1.1.1.35) were used as markers of glycolytic metabolism, global oxidative capacity (tricarboxylic cycle) and lipid beta-oxidation potential, respectively. Enzyme activities were measured on LM as described previously [38] and expressed as micro-moles of substrate degraded per minute per gram fresh muscle.

RNA preparation and gene expression data set

Total RNA was isolated from each of the 64 muscle samples as previously described in Voillet et al. [18]. After quality control and quantile normalization steps, 61 microarrays containing 44,368 probes were kept. The raw and normalized data are available in the Gene Expression Omnibus under the accession number GSE56301.

Integration of proteome network data

Network analysis was performed using a three step approach as illustrated in Figure 1 and at the two developmental time points separately (90 and 110 dg (32 individuals per condition)), resulting in two global networks. We chose to infer two networks to avoid the high gestational age effect already observed

in a previous study analyzing a transcriptomic dataset with the same individuals [18]. All analyses were performed using the R computing environment [39].

Figure 1 about here.

Inference of the proteome co-expression network

The first step consisted in inferring a proteome co-expression network using the PCIT (partial correlation and information theory) algorithm [40] (Figure 1A). PCIT has already been successfully used in transcriptomic analyses [41, 42]. In the present study, we performed PCIT on a proteomic data set. PCIT belongs to the family of weighted network algorithms and is based on the combination of the concept of partial correlation coefficient and the information theory to identify meaningful associations. Briefly, the co-expression arrangements for all triplets are compared, with all triplets being exhaustively explored. PCIT estimates the correlation for each pair of proteins taking the presence of a third protein into account. Significant correlations establish an edge in the reconstruction of the network. Even though PCIT is a soft-thresholding method, the number of selected edges was also adjusted according to the network density (around 5% for each network) to obtain readable networks. This first step of the analysis was performed with the R package PCIT [43].

Proteome network clustering

The second step consisted in finding communities within networks (Figure 1B). For each network, a spin-glass model was performed to optimize the modularity Q [44] and to cluster nodes [45]. The modularity Q is a measure of the quality of communities (or clusters) in a network: highly connected genes within each community and weakly connected genes between communities [44]. As already described in [46], a permutation test (100 permutations of edges corresponding to 100 random networks with the same degree distribution) was used to declare whether or not the clustering was significant.

Proteome network community analysis and relationship with a phenotype of interest

The third step consisted in the biological analysis of each community (or sub-network). GeneMANIA networks [47] were used to explore and confirm the relevance of the proposed communities. Gene ontology (GO) enrichment analysis using the GeneCodis webservice [48] was used to identify the biological functions

represented by each community (Figure 1C). For the significance of GO enrichment, multiple testing was controlled using the false discovery rate (FDR) approach [49].

The third step included highlighting some nodes in each community. As already reported in other studies [18,46], the betweenness of nodes within each community was analyzed (Figure 1C). A permutation test with 1,000 permutations was performed to check if the betweenness centrality was significant with respect to the node's degree. A significant result (FDR < 0.05) indicated a node was more central in the network than expected. Therefore, betweenness centrality is a good measure of the importance of the node in the network. A node with a high betweenness (or centrality) value has a marked influence on the structure of the network. All these analyses were performed using the R package igraph [50]. If no node with significant betweenness was found, the nodes with the highest betweenness were highlighted.

In an integrative strategy, phenotypic information was also added to the co-expression network. Links between sub-networks and biological phenotypes of interest were investigated (Figure 1C). To this end, methods coming from spatial statistics were used as described in [13,51]. First, the correlation between each protein expression community and the phenotype of interest was calculated. Second, a Moran's I was calculated to measure the spatial correlation between the sub-network structure and the phenotype of interest. Then, to check if the correlation between the biological phenotype and the sub-network was significant ($p < 0.05$), a permutation test with 1,000 node permutations was computed.

Network layout

All the graphs were laid out using the Force Atlas layout [52]. The degree (the number of adjacent edges) is indicated by the size of the node, and betweenness is indicated by the color of the node. The colors of the edges show whether the correlation between nodes is positive (in red) or negative (in blue). All these analyses were performed using Gephi software [53].

Gene-protein integration

Among the 89 unique proteins identified, 81 corresponding genes were available in the muscle transcriptome dataset [18]. When multiple probes mapped to the same gene, the highest differentially expressed probe according to the interaction between gestational time points and fetal genotypes was retained.

Transcriptomic and proteomic analyses were carried out using the same 64 muscle samples, but 60 fetuses were available in both proteomic and transcriptomic datasets.

To identify proteins with possible transcriptional regulation, for each protein within each fetal genotype, a Pearson's correlation was computed between mRNA (from the gene expression dataset [18]) and protein expression levels. A test was also run to assess if the correlation was significantly non-null ($p < 0.01$). Multiple testing was controlled using the FDR approach [49].

From a biological point of view, the Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity) software was used to check enrichment analysis (biological functions and canonical pathways), to construct bibliographic networks and regulation networks based on the identification of potential upstream regulators. A focus was on sub-network 2 at 110 dg, because of its relevant spatial correlation with two biological phenotypes of interest and a high number of proteins with a possible transcriptional regulation. Briefly, IPA constructed two separated networks. The first was based on bibliographic data in which the edges were obtained from biological links such as receptor-ligand interactions, enzyme activity on another protein, or a transcriptional factor activating the expression of targeted genes. IPA proposed the most probable network with an associated score. IPA also identified upstream regulators with a statistical likelihood of targeting some of the genes or proteins in the network. Finally, IPA generated a regulated network with the latest information. The proposed network (described in the Results section) was reconstructed from both IPA networks by integrating direct and regulated links. The PathDesigner function of IPA was used to draw a final graph with all previous information plus the information related to the correlation between transcripts and proteins when the information was available from the transcriptomic data.

Results

Analysis of MyHC polymorphism and muscle glycogen content

MyHC mRNA and protein levels and muscle glycogen content are listed in Table 1. All these biological phenotypes were differentially expressed according to the interaction between the two developmental time points (90 and 110 dg) and the four fetal genotypes (MS, LW, MSLW and LWMS), except slow (I + α-cardiac) MyHC at the protein level. The expression of IIb and extra-ocular MyHC were undetectable at 90 and 110 dg. In all four genotypes, embryonic and perinatal MyHC (mRNA and protein levels) decreased at the end of gestation, whereas fast (IIa + IIx) and slow (I + α-cardiac) MyHC increased during the maturation process. Embryonic MyHC was more highly expressed in LW than in MS fetuses at 90 dg, whereas a higher value of fast (IIa and IIx) MyHC was observed in MS than in LW at 110 dg. A high correlation between mRNA and protein levels was found for both embryonic and adult fast MyHC (Pearson's correlation between mRNA and protein levels equal to 0.95 and 0.93, respectively) (Figure 2).

Table 1 about here.

The profile of muscle glycogen content resembled that of adult fast MyHC between 90 and 110 dg with a 2.6 fold increase during the maturation process (Table 1). At 90 dg, MS and LWMS fetuses already exhibited higher muscle glycogen content than LW and MSLW fetuses. At 110 dg, levels reached 12.2% in MS and LWMS vs 10.5% and 9.5% in MSLW and LW, respectively ($p < 0.001$). Figure 2 shows the corresponding changes in muscle glycogen content and both embryonic and adult fast MyHC. Muscle glycogen content was particularly well correlated with embryonic MyHC at 90 dg and with adult fast (IIa + IIx) MyHC at 110 dg.

Figure 2 about here.

Because MS piglets are known to be more mature than LW piglets at birth [15], embryonic MyHC, fast (IIa and IIx) MyHC and muscle glycogen content are likely good descriptors of muscle physiological maturity in pig neonates. These biological phenotypes were consequently used in the proteomic and transcriptomic analyses to help identify proteins and genes potentially involved in the muscle maturation process in late fetal stages.

Proteomic analysis

Multiple statistical analyses to select potentially relevant protein spots

As described in Materials and Methods, several statistical methods were conducted to obtain a large spectrum of proteins related to the experimental design. One hundred seventy nine spots were selected among the 1,025 spots matched between gels. After protein identification, 120 spots, corresponding to 89 unique proteins, were successfully identified by LC-MS/MS and 18 proteins exhibited different isoforms. To facilitate data interpretation and computation, protein isoforms were filtered to exclude redundant isoforms (i.e. highly positively correlated isoforms). When multiple highly positively correlated isoforms (Pearson's correlation > 0.9) mapped to the same protein, only the most differentially expressed isoform for the interaction between developmental time points and fetal genotypes was retained. Isoforms without a high positive correlation were kept (Pearson's correlation < 0.9). Finally, a total of 113 proteins were retained and used for further analyses (Additional File 5).

The distribution of the 113 identified spots among the different statistical analyses is shown as a Venn diagram in Figure 3. Seventy-one spots were significantly differentially expressed of which 13 were affected by the interaction and 58 by the additive model between gestational time points and fetal genotypes. Using RF and sPLS-DA analyses to obtain spots that distinguished gestational stages and/or fetal genotypes, 55 spots were selected. Then, 62 spots were also selected using several multivariate methods to obtain spots correlated with the biological phenotypes of interest (MyHC polymorphism and glycogen content). Altogether, a large number of spots overlapped between the multivariate and discriminant analyses, so that 179 spots were finally selected.After spot identification by mass spectrometry and some isoform removing, 113 proteins were identified.

Figure 3 about here.

The classification of the 113 identified proteins according to the biological axes is presented as a level plot in Figure 4. Proteins involved in energy metabolism processes, such as glucose metabolism, gluconeogenesis, oxidation-reduction activity and mitochondrion, were mostly over-expressed at 110 dg in all genotypes. The figure also shows that oxidation-reduction related to mitochondria at 110 dg increased in the order LW $<$ MSLW $<$ LWMS $<$ MS. On the other hand, proteins involved in muscle development, such as system development, actin skeleton, muscle filament sliding, cytoskeleton and mRNA metabolic process, were

mostly over-expressed at 90 dg in all genotypes. All enriched biological functions and cellular components are not shown in this figure, as we chose to highlight only important GO terms with a high number of proteins. For that reason, some proteins do not belong to any of the biological processes shown.

Figure 4 about here.

Proteome sub-network analysis

Network inference was performed at each developmental time point according to the three-step inference method presented in Figure 1. The largest connected component of the d90 and d110-proteome networks are presented in Additional File 6 and characteristics of these networks (degree, betweenness and clustering) are summarized in Additional Files 7 and 8, respectively.

d90-proteome network analysis

The largest connected component of the d90-proteome network was composed of 94 nodes and 314 edges (density of 7.2%) and followed a scale-free topology denoting a non-random organization of the network [10]. After node clustering, five sub-networks were obtained (Table 2 and Figure 5A). All five sub-networks displayed between 16-21 nodes and 26-55 edges. Additional File 9 displays all five sub-networks, and Additional File 10 shows the enriched Gene Ontology (GO) terms for each sub-network. In several sub-networks, some isoforms for the same protein were linked because only one developmental time point was used for the network inference. Figure 5A shows that sub-networks 1, 2 and 4 had a high number of edges in common, and shared a GO term corresponding to muscle filaments. Notably, these three sub-networks were also those that were significantly and spatially correlated with the three biological phenotypes of interest (Table 2). Figure 6 shows the three sub-networks (sub-networks 1, 2 and 4) were significantly and spatially correlated with the biological phenotypes of interest. Sub-network 1 was correlated with glycogen and embryonic MyHC, sub-network 2 to glycogen and adult fast (IIa + IIb + IIx) MyHC, whereas sub-network 4 was only correlated with embryonic MyHC.

Table 2 about here.

Figure 5 about here.

Figure 6 about here.

According to GO functional enrichment analysis, each sub-network was related to several biological functions (Table 2): sub-network 1 was mainly involved in the myofibril (GO cellular component (CC)), glycolysis (GO biological process (BP)) and mitochondrion (CC), sub-network 2 in actin filament (CC), gluconeogenesis (BP) and mitochondrion (CC), and sub-network 4 in muscle filament sliding (BP), gluconeogenesis (BP) and cell cycle checkpoint (BP). Sub-network 3 was mainly related to the respiratory electron transport chain (BP), cell redox homeostasis (BP) and mitochondrion (CC), whereas sub-network 5 was related to the creatine metabolic process (BP) and sarcolemma (CC). It is important to note that some identical GO terms were enriched in several sub-networks because a large number of the proteins identified exhibited different isoforms and these isoforms could be present in different sub-networks. In addition, the proteins we identified could also be involved in several metabolic functions.

To go deeper into the biological interpretation of the clusters, we chose to highlight the three sub-networks correlated with the biological phenotypes of interest (Figure 6 and Table 2). Sub-network 1 showed a significant correlation with muscle glycogen content and embryonic MyHC. The enriched GO biological processes, such as glycolysis, glucose metabolic process and gluconeogenesis, were in agreement with the glycogen correlation. FETUB (Fetuin-B, a member of the cysteine protease inhibitor family - involved in the negative regulation of endopeptidase activity) and OXCT1 (Succinyl-CoA:3-ketoacid-coenzyme A transferase 1, mitochondrial - involved in ketone body catabolic process) exhibited the highest degree of this sub-network. In addition, FETUB exhibited the highest betweenness and was more highly expressed in LW than in MS fetuses. Muscle filament sliding was also one of the enriched GO terms describing this sub-network. In sub-network 2, one isoform (isoform 2) of PSMC5 (Proteasome 26S Regulatory Subunit, ATP-dependent degradation of ubiquitinated proteins, 5) had significantly high betweenness and the highest degree of this sub-network. This isoform of PSMC5 was more highly expressed in MS than in LW fetuses (Figure 7). Three isoforms (isoforms 1, 2 and 4) of GPD1 (Glycerol-3-phosphate dehydrogenase 1, cytoplasmic, involved in the glycerol phosphate shuttle) were also identified in this sub-network. Interestingly, these three isoforms were more highly expressed in MS than in LW fetuses, whereas GPD1 isoform 3 was less expressed in MS than in other genotypes (Figure 7). Muscle filament sliding was also identified as a significant GO term to characterize sub-network 2. Sub-network 4 was correlated with the embryonic MyHC. One isoform (isoform 2) of TNNT3 (Troponin T type 3) had

significantly high betweenness and the highest degree of this sub-network. The enriched GO biological processes, e.g. muscle filament sliding and development and cell cycle checkpoint, were in agreement with the values of the phenotypic correlation. Two other TNNT3 isoforms (isoform 1 and 3) were also present in this sub-network. It is noteworthy that a common GO term applied to sub-clusters 1, 2 and 4 related to muscle filaments, which could denote a strong involvement of muscle filaments in the maturational process at 90 dg, as previously highlighted on the level plot in Figure 4.

Figure 7 about here.

d110-proteome network analysis

The largest connected component of the d110-proteome network was composed of 86 nodes and 313 edges (density of 8.6%) and had scale-free topology denoting a non-random organization as observed at 90 dg. We obtained six clusters (Table 3 and Figure 5B). Additional File 11 shows all six sub-networks and Additional File 12 shows the enriched GO terms for each sub-network. Among these six sub-networks, five sub-networks displayed more than 10 nodes and 19 edges. Four sub-networks were spatially correlated with muscle glycogen content (sub-networks 3, 4, 5 and 6), whereas three sub-networks were spatially correlated with the embryonic MyHC (sub-networks 2, 3 and 4) or the adult fast MyHC (sub-networks 2, 3 and 5).

Table 3 about here.

GO term enrichment analysis was performed for each sub-network (Table 3): (i) sub-network 1 (only five nodes) was involved in the tricarboxylic acid cycle (BP) and mitochondrion (CC), (ii) sub-network 2 consisted in 24 nodes and 65 edges and was primarily involved in the tricarboxylic acid cycle (BP), respiratory electron transport chain (BP), glucose metabolic process (BP) and muscle filament sliding (BP), (iii) sub-network 3 was involved in gluconeogenesis (BP) and glycolysis (BP), (iv) sub-network 4 in the mRNA metabolic process (BP) and in the proteasome complex (CC), (v) sub-network 5 in muscle development (BP) and also lipid metabolism (BP), and (vi) sub-network 6 was mainly involved in gluconeogenesis (BP), ATP catabolic/anabolic process activity (BP) and mitochondrion (CC).

As previously, we chose to highlight the five sub-networks correlated with a biological phenotype of interest (Figure 10). Sub-network 2 was correlated with both adult fast and embryonic MyHC and had

the highest number of nodes (24) and edges (65). This sub-network was primarily characterized by GO terms corresponding to mitochondrial oxidative metabolism (Additional File 12). ATP5A1 (ATP synthase subunit alpha, mitochondrial) showed significant high betweenness and CKMT2 (Creatine kinase, mitochondrial 2) exhibited the highest degree of the sub-network. Both CKMT2 and ATP5A1 were more highly expressed in MS than in LW fetuses (Figure 7). These proteomic results were validated for CKMT2 by western blotting carried out on the 64 fetuses (Figure 8). Briefly, CKMT2 increased dramatically between 90 and 110 dg and was greater in MSMS fetuses at 110 dg, whereas MSLW and LWMS were intermediate between LWLW and MSMS fetuses. This was also supported by enzyme activities with a greater increase in activities of CS and HAD between 90 and 110 dg in MSMS fetuses (+283 and +122% for CS and HAD, respectively) than in LWLW fetuses (+172% and +87%, respectively) (Figure 9). Moreover, the differential increase between MSMS and LWLW fetuses was more important for CS than HAD activities. Besides that, the activity of LDH did not differ between MSMS and LWLW fetuses at 90 dg, remained stable in LWLW up to 110 dg and increased moderately (+31%) in MSMS fetuses (Figure 9). Sub-network 3 was significantly correlated with all three phenotypes of interest and was mainly involved in glucose metabolic process, gluconeogenesis and glycolysis occurring in the cytoplasmic compartment. GSN (Gelsolin - involved in actin filament reorganization) was the protein with the highest betweenness, but also the highest degree of this sub-network with PDIA3 also known as GRP58 (protein disulfide-isomerase A3 - involved in protein folding). Sub-network 4 was correlated with the embryonic MyHC and muscle glycogen content and mostly involved in the proteasome complex and the mRNA metabolic process. An isoform (isoform 1) of PSMC5 had the highest betweenness. This protein and DDAH1 (Dimethylarginine dimethylaminohydrolase - involved in arginine metabolic process) had the highest degree of this sub-network. Muscle glycogen content and adult fast MyHC were the two biological phenotypes correlated with sub-network 5. PGAM1 (Phosphoglycerate Mutase 1 - involved in the glycolytic process) was the protein with the highest degree, whereas one isoform (isoform 2) of LDB3 (LIM domain binding 3 - involved in sarcomere organization) had the highest betweenness. Sub-network 6 was significantly correlated with muscle glycogen content, which is consistent with the enriched GO terms (*e.g.* gluconeogenesis and ATP catabolic/anabolic process). An isoform (isoform 2) of PSCM5 had the highest betweenness and the highest degree of this sub-network. Thus, PSMC5 isoform 2 had a major influence on the structure of sub-network 6 at 110 dg and sub-network 2 at 90 dg, whereas PSCM5 isoform 1 played a key role in sub-network 4 at 110 dg. Two isoforms (isoform 2 and 4) of GPD1 were

also present in sub-network 6 and under-expressed in LW fetuses (Figure 7).

Figure 8 about here.

Figure 9 about here.

Figure 10 about here.

Identification of mRNAs correlated with proteome networks

For each protein (with a corresponding gene in the transcriptomic dataset) within each extreme fetal genotype, a Pearson's correlation was computed between mRNA and protein expression levels in order to identify proteins with possible transcriptional regulation. The transcriptomic dataset was previously used in [18] and 60 fetuses were represented in both the transcriptomic and proteomic datasets. figure 11 shows the correlation between transcriptome and proteome in MS (x axis) and LW (y axis) genotypes. Among the 81 proteins available in the transcriptomic dataset, 31 proteins exhibited a significant positive correlation between mRNA and protein expression levels in both extreme fetal genotypes, 8 proteins in LW only (mostly involved in muscle filament sliding and located in cytosol) and 13 proteins in MS only (with no specific biological enrichment) (Figure 11 and Additional File 13). On the other hand, one protein and its associated isoform (TNNT3 isoform 6) was also found to have a significant negative correlation between mRNA and protein expression levels in both extreme genotypes, two proteins in LW only (TNNT3 isoform 1 and LDB3 isoform 1) and two proteins (TNNT3 isoform 7 and GSN isoform 1) in MS only (Additional File 13).

figure 11 about here.

For a more detailed biological interpretation, we chose to highlight the 31 proteins with a significant positive correlation between mRNA and protein expression in both MS and LW pure fetuses. Sub-network 2 at 110 dg, already highlighted because of its relevant spatial correlation with two biological phenotypes of interest (adult fast and embryonic MyHC), was particularly interesting. Thanks to the correlation analysis, we showed that 10 nodes (seven belonging to the mitochondrion cellular component) among the 24 nodes that comprised this sub-network exhibited a positive correlation between mRNA and protein expression and were likely transcriptionally regulated (Additional File 13), especially CKMT2 and ATP5A1. An IPA analysis of this sub-network (using all nodes) identified possible transcription factors (TFs) affecting

these proteins, in particular PPARGC1A (peroxisome proliferator-activated receptor gamma coactivator 1-alpha) with an effect on CKM (creatine kinase, muscle, cytoplasm), ATP5A1, CKMT2 and ACADVL (very long-chain specific acyl-CoA dehydrogenase, mitochondrial - involved in fatty acid beta-oxidation) (Figure 12). The PPARGC1A gene (available in the transcriptomic dataset) was over-expressed at 110 dg compared to 90 dg. It is interesting to note that at 90 dg, the expression of PPARGC1A was lower in LW than in MS. This difference between genotypes, which was only visible at 90 dg, suggests that the increase in PPARGC1A expression was delayed at 90 dg in LW, which could have subsequently reduced protein expression of CKM, ATP5A1, CKMT2 and ACADVL at 110 dg in LW vs. MS fetuses. ESR1 (Estrogen receptor alpha) was also seen to affect targets mostly involved in proliferation and differentiation of muscle cells such as MACROD1 (MACRO Domain Containing 1), PDIA3 (isoform 2), SEPT2 (Septin 2), ANXA2 (Annexin A2) and GSN (Figure 12). The ESR1 gene (also available in the transcriptomic dataset) was over-expressed at 110 dg with a higher expression in MS than in LW at both 90 and 110 dg. IPA identified KCNJ11, a subunit of the ATP-sensitive K+ channel (KATP), as a regulator of many proteins involved in energy metabolism that were up-regulated between 90 and 110 dg and over-expressed in MS compared to LW at 110 dg. Expression of KCNJ11 at the mRNA level increased between 90 and 110 dg but did not differ between genotypes. However, KCNJ11 is not a transcription factor but is involved in the regulation of K+ ions in response to the ATP/ADP ratio [54] and cannot be considered as a potential upstream transcriptional regulator of muscle maturity. Finally, IPA analysis identified several myogenic factors whose expression decreased between 90 and 110 dg, and was lower in MS than in LW at 110 dg (see insert in Figure 12), in accordance with better maturity of MS at birth.

figure 12 about here.

Discussion

Embryonic MyHC, adult fast (IIa + IIx) MyHC and muscle glycogen content are good descriptors of neonatal muscle maturity

In the present experiment, the adult fast IIa and IIx MyHC were already expressed before birth, whereas no IIb was detected, which is in accordance with results obtained by [55]. In a previous study, we also found that α-cardiac MyHC was transitorily expressed in pig skeletal muscle shortly after birth [29], and the present results show that the α-cardiac MyHC is already expressed in late gestation. In all genotypes, embryonic and perinatal MyHC (mRNA and protein levels) decreased at the end of gestation, whereas fast (IIa + IIx) and slow (I + α-cardiac) MyHC increased during the maturation process. These results are in agreement with those of a previous review [56] which exhaustively describes the ontogenesis of skeletal muscle in farm species. In addition, because of a high correlation between mRNA and protein expression (Figure 2), our results suggest transcriptional regulation of MyHC expression during the maturation process. Lefaucheur et al. [57] already showed that expression of adult MyHC is very similar at the mRNA and the protein levels in pig skeletal muscle at 60 kg body weight. This transcriptional regulation has also been observed for other MyHC isoforms and is in accordance with previous results in skeletal muscle in other species [58].

The maturation of skeletal muscle contractile and metabolic properties has been reported to influence piglet maturity at birth [59, 60]. Among the MyHC phenotypes, embryonic and adult fast (IIa + IIx) MyHC appear to be good markers of muscle maturity at 90 and 110 dg, respectively (Table 1). Indeed, because embryonic MyHC was highly expressed at 90 dg but decreased drastically in late gestation, it is likely a good marker of muscle immaturity at 90 dg (easily measurable because highly expressed) and its higher expression in LW than MS points to a lower maturity of LW than MS fetuses at 90 dg. Conversely, fast (IIa + IIx) MyHC was highly expressed at 110 dg and increased dramatically in late gestation, making it a potentially good descriptor of muscle maturity around birth, and its higher expression in MS than in LW fetuses points to higher maturity of MS than LW fetuses at 110 dg. Interestingly, the hybrid fetuses were mostly intermediate between pure genotypes. In pigs, the adult fast IIa, IIx and IIb MyHC progressively replace the developmental MyHC (embryonic and perinatal MyHC) in late gestation and early after birth [56]. Because MS neonates are known to be more mature than LW, with a lower rate of

mortality at birth [15], the lower expression of embryonic MyHC at 90 dg and the higher expression of adult fast (IIa and IIx) MyHC at 110 dg are representative of a better muscle maturity in MS than in LW fetuses in late gestation, the hybrid fetuses being in an intermediate position.

In addition, good correspondence was found between the increased expression of adult fast (IIa + IIx) MyHC and muscle glycogen content between 90 and 110 dg (Figure 2). Energy reserves, e.g. glycogen and lipids, must be maximal in the neonatal period because piglets are not able to oxidize protein efficiently before 5-7 days of life [59]. In pig neonates, muscle glycogen content is very high (about 9% of fresh muscle vs. 1% in growing pigs [59]) and plays a key role in whole body glycogen storage (89% of total body glycogen at birth [59]) and thermoregulation because neonatal pigs are devoid of brown adipose tissue and only have a small amount of adipose tissue (about 1.5% of body weight). The animal's energy requirement is maximum in the neonatal period to promote locomotion, thermoregulation and growth. Therefore, a high level of muscle glycogen at birth is likely involved in better neonatal maturity and piglet survival. The higher muscle glycogen content found in MS fetuses in the present study is in agreement with the better maturity previously reported in MS vs. LW neonates [15]. Interestingly, a difference between the two extreme breeds in expression of genes involved in glycogen metabolic processes was already observed during the same fetal period in the same individuals [18]. Thus, some genes were over-expressed at 110 dg in MS only, such as *PCK2* (phosphoenolpyruvate carboxykinase 2, mitochondrial also known as PEPCK 2) likely involved in gluconeogenesis from precursors derived from the citric acid cycle [18]. Therefore, the significant difference we observed in muscle glycogen content between MS and LW can be, at least partly, explained by these differences in the transcriptome. Interestingly, MS fetuses also exhibited a greater PYGM (muscle glycogen phosphorylase) expression at the mRNA level than LW at 110 dg, which denotes a higher capacity to use glycogen as an energy source in MS fetuses. Skeletal muscle appears to play a key role in glycogen storage and oxidative metabolism around birth. We hypothesize that piglets with high value for survival, like MS, have a higher ability to maintain glucose levels during and after farrowing, and are more able to maintain body temperature.

Taken together, the present data showed that embryonic MyHC, adult fast (IIa + IIx) MyHC and muscle glycogen content are good descriptors of muscle maturity in pig fetuses during late gestation, demonstrate that MS fetuses are more mature than LW fetuses, and that hybrid fetuses are intermediate between pure MS and LW breeds. All these three biological phenotypes were used in the further data integration

analysis to help find new biological markers of neonatal maturity and to advance our understanding of the underlying mechanisms.

Mitochondrial oxidative metabolism is a key determinant of neonatal muscle maturity

Oxidative metabolism tends to increase during late gestation in all farm species making it an increasingly important source of energy during fetal life [56]. To identify the biological processes underlying muscular maturity, the GO terms classification was performed on the 89 unique identified proteins. A large number of proteins associated with metabolic processes were identified. However, 2D gel electrophoresis is known to reveal a limited collection of highly abundant and soluble proteins [61] covering a limited number of cell functions, mainly dealing with cell structure and metabolism, which is why proteomic and transcriptomic approaches are not always comparable [62].

Network analysis was performed and computed at the two developmental time points to identify proteins with a relevant role in the network and to identify the biological processes. It is important to note that, at both 90 and 110 dg, a large number of biological processes and cellular components overlapped between sub-networks and, within each sub-network, the enriched biological analysis often highlighted a mixture of GO terms. Indeed, it is sometimes difficult to assign a single biological function to a protein because they often play several roles in interdependent metabolic pathways. Several isoforms were also found in the proteomic dataset, and were diversely intercorrelated meaning two isoforms of the same protein could be found in two different sub-networks.

The GO terms over-expressed at 90 dg were mostly involved in the actin cytoskeleton, muscle development and mRNA processing (Figure 4). As already highlighted in a previous transcriptomic study using the same individuals [18], these results are consistent with the second phase of myofiber genesis known to occur between 55 and 90 dg in pigs [56]. The total number of muscle fibers increases up to approximately 90 dg and further muscle development mostly occurs through hypertrophy and maturation of existing muscle fibers [56]. In addition, it has already been reported that the total number of myofibers is lower in MS than LW [57] which helps explain the lower postnatal muscle growth capacity of MS than LW pigs. Among the proteome sub-networks that were correlated with the biological phenotypes of interest

(sub-networks 1, 2 and 4), all were characterized by GO terms involved in muscle filament development and sliding, underlining the importance of filamentous proteins and likely of the cytoskeleton at this stage. Notably, sub-network 2 also contained three isoforms of GPD1 (isoforms 1, 2 and 4) that were over-expressed in MS fetuses (Figure 7). GPD1 is a cytoplasmic enzyme involved in the glycerol phosphate shuttle that can transfer cytoplasmic glycolytic reducing NADH equivalents to mitochondrial FADH2 at complex 2, thus providing ATP to the cell through the respiratory chain from complex 2 [63]. The higher expression of these three GPD1 isoforms in MS at 90 dg could suggest increased mitochondrial oxidation of cytosolic NADH in MS, which already contributes to the advanced maturity of MS fetuses. Surprisingly, sub-networks 3 and 5, which were mostly involved in energy metabolism, had no significant correlation with MyHC and muscle glycogen content at 90 dg. On the whole, our data suggest that muscle immaturity at 90 dg is primarily related to the high proportion of cytoskeletal proteins and proteins involved in myofibril assembly, even though some metabolic enzymes such as GPD1 could also be positively correlated with the maturation process. In accordance with previous data in different farm species [56], a dramatic increase in muscle oxidative metabolism occurred between 90 and 110 dg as shown by the proteomic, western blotting and enzymatic data in the present experiment.

At 110 dg, the most striking differences between genotypes concerned the mitochondria cellular component, in particular the mitochondrial oxidation/reduction molecular function (Figure 4). Moreover, the proteome sub-network analysis identified gluconeogenesis and glycolysis as important co-expressed pathways that could explain the higher muscle glycogen content in MS than in LW at 110 dg. Analysis of the proteome sub-network identified five sub-networks that were correlated with the biological phenotypes of interest (sub-networks 2, 3, 4, 5 and 6). Most were characterized by GO terms primarily involved in energy metabolism (sub-networks 2, 3, 5 and 6), whereas GO terms dealt mostly with the proteasome complex and mRNA processing in sub-network 4 with PSMC5 (isoform 1) as the most important node in the cluster. The most relevant sub-network was sub-network 2 with 24 nodes and 65 edges. This sub-network was highly correlated with adult fast and embryonic MyHC, and its GO terms primarily concerned mitochondrial energy metabolism. Interestingly, CKMT2 and ATP5A1 were identified as important nodes in this cluster which could be physiologically relevant for the production of energy such as ATP. Indeed, CKMT2 is a creatine kinase and is responsible for the transfer of high energy phosphate from the mitochondria to the cytosolic compartment, and at the same time for returning ADP to the respiratory system, thereby

stimulating oxidative phosphorylation. Moreover, cytoplasmic muscle CKM was also over-expressed in MS at 110 dg, suggesting that the energy metabolism of muscle contraction was higher in MS than LW. ATP5A1 encodes a subunit of the mitochondrial ATP synthase that converts the mitochondrial electrochemical H^+ gradient to ATP production, thereby supplying ATP to the muscle. CKMT2 and ATP5A1 are likely good biological markers of muscle maturity at 110 dg (Figure 7 and Figure 8). The greater expression of CKMT2 and ATP5A1 in MS at 110 dg indicates that MS fetuses possess greater oxidative capacity, which was supported by the greater activities of CS (tricarboxylic acid cycle) and HAD (lipid beta-oxidation) enzymes (Figure 9), and in accordance with previous results showing greater enzyme activities of CS and HAD in MS than LW muscle at birth [64]. Sub-network 3 was the second most important cluster with 19 nodes and 46 edges. It was highly correlated with glycogen content as well as embryonic and adult fast MyHC, and its GO terms mostly concerned glycolytic energy metabolism and gluconeogenesis. Sub-network 6 was highly correlated with muscle glycogen content and logically, one of its GO terms corresponded to gluconeogenesis. Moreover, two GPD1 isoforms (isoforms 2 and 4) were also present in sub-network 6 and under-expressed in LW fetuses, as already observed at 90 dg (Figure 7). Notably, expression of the four GPD1 isoforms was strongly influenced by the genotype with no effect of age, and no interaction between genotype and development time points was observed (Additional File 5). This suggests that GPD1 is genetically determined but not related to maturity as influenced by age. At the mRNA level, as already demonstrated in [18], GPD1 gene expression was lower in LW than MS at 90 dg and did not increase in LW during the maturation process, whereas it did increase in MS. Therefore, the gestational time point only had an effect on GPD1 at the mRNA level in MS. The greater GPD1 expression in MS than LW at 110 dg could be physiologically relevant through the transfer cytoplasmic NADH to mitochondrial FADH2 (glycerol phosphate shuttle), which could contribute to the production of energy through the oxidative phosphorylation in the respiratory chain. Interestingly, the mRNA expression level of LDHA, which catalyzes the conversion of L-lactate and NAD to pyruvate and NADH in the final step of anaerobic glycolysis was greater in MS than LW fetuses [18] and the decrease in blood lactate concentration between the fetal vein and fetal artery was greater in MS than LW fetuses (-0.18 +/- 0.09 vs 0.05 +/- 0.20 mmol/L, respectively, P = 0.01), suggesting a greater use of lactate, mostly produced by the placenta, as an energy source in MS than LW fetuses. The enzyme activity of LDH was significantly greater in MSMS than LWLW fetuses (+17%, P = 0.04, Figure 9). However, LDH is a reversible enzyme which can transform either pyruvate to lactate (+ NAD+) or lactate to pyruvate (+ NADH) depending

on its isozymic composition and substrate availability, making it necessary to perform further work to better understand the real metabolic fluxes through the LDH reaction. Because LDH activity dramatically increases from birth to 15-20 days of age [65], the greater LDH activity in MSMS than LWLW fetuses at 110 dg could also be related to better maturity of MSMS piglets at birth. Altogether, the present data show that muscle energy metabolism, in particular mitochondrial energy metabolism and muscle glycogen storage, increased dramatically between 90 and 110 dg and were higher in MS than in LW fetuses at 110 dg, meaning that the mitochondrial oxidation/reduction process and glycogen storage, likely involving the mitochondrial PEPCK (PCK2), play a crucial role in the late fetal muscle maturation process.

Biological data integration identified PPARGC1A as potential upstream transcriptional regulator of muscle maturity at birth

To gain further insight into the molecular machinery underlying the muscle maturation process in pigs, we undertook a guided and integrated analysis of the transcriptomic and proteomic datasets. Among the 89 unique proteins identified, 81 were available in the transcriptomic dataset. Using Pearson's correlations, mRNA and protein expression levels were analyzed to find proteins with possible transcriptional regulation. Thirty-one proteins were identified with a significant positive correlation between mRNA and protein expression levels in both extreme fetal genotypes (Pearson's correlation $|r| > 0.7$) (Figure 11). A number of transcriptome and proteome data integration studies have already been reported *e.g.* in cell lines [66–68], plants [69], and mammal models [6, 70–73]. The correlation between mRNA and protein abundances in the cell has been reported to be notoriously poor. At 110 dg, sub-network 2 was composed of 10 nodes (out of 24) with possible transcriptional regulation in both MS and LW. Proteins belonging to this sub-network generally exhibited similar expression profiles at 110 dg, were mostly located in the same cellular component and may possibly be regulated by common transcription factors (TFs). Based on a bibliographic network computed using IPA software (using the upstream regulator function), we were able to find possible upstream TFs. PPARGC1A was one of the TFs identified. It was a regulator of CKMT2, ACADVL and ATP5A1 (Figure 12). CKMT2 has already been identified as an important node (the highest betweenness for CKMT2 at 110 dg) and as a likely good biological marker of muscle maturity at 110 dg.

PPARGC1A (also known as PGC1-alpha) is a transcriptional coactivator that controls the expression

of many genes through a whole range of nuclear hormone receptors and other TFs [74]. Moreover, the activity of PPARGC1A is regulated by phosphorylation and deacetylation through the coaction of two upstream metabolic sensors of energy deficiency: AMPK and SIRT1, respectively [75]. PPARGC1A is abundant in muscle [76] where it is involved in several biological functions such as mitochondrial biogenesis, and oxidative metabolism, which play a key role in ATP production and the adaptation of muscle to exercise and exposure to cold [77]. It has been shown to drive the formation of slow twitch fibers [78] and to be more highly expressed in oxidative myofibers [75]. Therefore, together with AMPK and SIRT1, PPARGC1A is involved in carbohydrate and lipid metabolisms to maintain energy homeostasis. In post-natal growing pigs, PPARGC1A is more highly expressed in LM of Erhualian than LW pigs, in accordance with a higher proportion of oxidative fibers in Erhualian pigs [79], as well as in MS pigs [57]. Voillet et al. [18] showed that PPARGC1A gene expression was differentially expressed depending on the interaction between fetal genotypes and gestational time points, with higher expression in MS than LW at 90 dg. Therefore, our results suggest that PPARGC1A could have a precocious effect on the subsequent greater increase in expression of genes such as CKMT2, ATP5A1, ACADVL and CKM in MS than LW at 110 dg. Mitochondrial CKMT2 and cytoplasmic CKM are both involved in the creatine metabolic process and are known to play a key role in the transfer of energy within the muscle fiber for muscle contraction and thermogenesis. Interestingly, over-expression of PPARGC1A in the skeletal muscle of transgenic mice has been shown to increase glycogen synthesis and storage [80], which could also explain the greater glycogen content observed in MS compared to LW fetuses at 110 dg in the present experiment. In addition, greater variability of PPARGC1A expression in LW than in MS was also observed at 110 dg. Therefore, PPARGC1A could be a relevant upstream regulator involved in the accelerated muscle maturation observed in MS compared to LW in late gestation.

Cell culture experiments have shown that PPARGC1A is a coactivator of ESR1 [81]. Interestingly, ESR1 activation has also been reported to increase the expression of PPARGC1A, NRF1, IRS1 and GLUT4 in skeletal muscle of ovariectomized mice fed a high fat diet [82]. In our transcriptomic study, ESR1 expression increased dramatically between 90 and 110 dg and was greater in MS than LW fetuses at both stages (Figure 12). In sub-network 2 at 110 dg, ESR1 was found to be a regulator of several proteins mostly involved in protein folding (PDIA) and in cytoskeleton (SEPT2 and GSN). This estrogen receptor has been reported to be involved in estrogen-mediated regulation of substrate metabolism [83] and lower

mRNA expression has been observed in the adipose tissue of obese compared to lean women [84]. In ESR1 knock out mice, body weight was about 30% higher than in wild-type mice: mice were obese and their oxidative metabolism was impaired [85]. In our study, under-expression of ESR1 in LW was accompanied by reduced oxidative metabolism but not by increased fatness (LW leaner than MS). Altogether, ESR1 increased between 90 and 110 dg, and its over-expression in MS compared to LW could be involved, in interaction with PPARGC1A, in the improved maturity of MS compared to LW animals at birth.

Conclusion

In this study, we developed an innovative strategy combining state-of-the-art statistical and computational methods to integrate multi-omics datasets to gain a deeper insight into the late fetal maturation process. Three biological phenotypes of interest (i.e. adult fast and embryonic MyHC and muscle glycogen content) were identified as good descriptors of maturity. Some proteins, involved in mitochondrial oxidative metabolism and with a possible transcriptional regulation, were also emphasized as possible biomarkers of the maturation process. As one of the main sources of energy during late fetal life, muscle oxidative metabolism is very important at birth [56], and likely strongly influences muscle and whole body metabolic maturity at the time of birth. In pigs, a higher rate of death at birth has already been observed in genotypes with a lower percentage of oxidative fibers [15,57]. In recent decades, the pig industry has been focusing on selection for rapid production of lean meat, low back-fat thickness and a rapid growth rate, thus influencing muscle fiber properties [86]. More precisely, a selection for lean tissue is more associated with muscles containing a low percentage of oxidative myofibers and a high percentage of large glycolytic myofibers [87]. This kind of the selection could have an effect on important genes, such as PPARGC1A and ESR1, with lower expression in skeletal muscle, which could contribute to an alteration or delay of mitochondrial gene/protein expression in late gestation and lead to muscle metabolic immaturity at birth.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We are grateful to P. Ecolan and S. Tacher for 2D gel analyses, C. Tréfeu and S. Daré for RT-PCR analyses and glycogen determination, respectively, D. Viala for LC-MS/MS protein identification and Annie Vincent for validation of CKMT2 differential expression by western blot. We also wish to thank D. Goodfellow for English revision. We also thank Kim-Anh Lê Cao for her relevant comments.

Funding

Animal Genetics Division (INRA, http://www.ga.inra.fr/en/), Animal Physiology and Livestock Systems (INRA, http://www.phase.inra.fr/en/) and *Région Languedoc Roussillon Midi Pyrénées* (http://www.regionlrmp.fr/) (to V.V.); French 'Agence Nationale de la Recherche' Porcinet grant [ANR-09-GENM005, http://www.agence-nationale-recherche.fr/]. Funding for open access: INRA, UMR1388 Génétique, Physiologie et Systèmes d'Élevage, Castanet-Tolosan, F-31326, France.

Author contributions

LLi, LC and LL conceived and designed the study. YB supervised the performance testing, from animal production to biological sampling. VV analyzed the expression datasets supervised by MSC, LL and LLi. VV drafted the manuscript with help of MSC, LL and LLi. LLi and LL supervised the project. All authors read and approved the final manuscript.

References

1. Metzker, M. L. (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31-46.

2. Ozsolak, F. and Milos, P. M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12, 87-88.

3. Altelaar, A. F. M., Munoz, J., and Heck, A. J. R. (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* 14, 35-48.

4. Shulaev, V. (2006) Metabolomics technology and bioinformatics. *Brief Bioinform* 7, 128-139.

5. Haider, S. and Pal, R. (2013) Integrated analysis of transcriptomic and proteomic data. *Curr Genomics* 14, 91-110.

6. Cox, B., Kislinger, T., and Emili, A. (2005) Integrating gene and protein expression data: pattern analysis and profile mining. *Methods* 35, 303-314.

7. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 27, 117-124.

8. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2013) Global quantification of mammalian gene expression control. *Nature* 473, 337-342.

9. Ritchie, M., Holzinger, E., Li, R., Pendergrass, S., and Kim, D. (2015) Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* 16, 85-97.

10. Barabàsi, A. and Oltvai, Z. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5, 101-113.

11. Luscombe, N., Babu, M., Yu, H., Snyder, M., Teichmann, S., and Gerstein, M. (2004) Genomics analysis of regulatory network dynamics reveals large topological changes. *Nature* 431, 308-312.

12. Mitra, K., Carvunis, A., Ramesh, S., and Ideker, T. (2013) Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* 14, 719-732.

13. Villa-Vialaneix, N., Liaubet, L., Laurent, T., Cherel, P., Gamot, A., and SanCristobal, M. (2013) The structure of a gene co-expression network reveals biological functions underlying eqtls. *PloS ONE* 8, e60045.

14. Canario, L., Cantoni, E., Le Bihan, E., Caritez, J. C., Billon, Y., Bidanel, J. P., and Foulley, J. L. (2006) Between-breed variability of stillbirth and its relationship with sow and piglet characteristics. *J Anim Sci* 84, 3185-3196.

15. Canario, L., Pere, M. C., Tribout, T., Thomas, F., David, C., Gogué, J. M., Herpin, P., Bidanel, J. P., and Le dividich, J. (2007) Estimation of genetic trends from 1977 to 1998 of body composition and physiological state of large white pigs at birth. *Animal* 1, 1409-1413.

16. Leenhouwers, J., Knol, E., de Groot, P., Vos, H., and van der Lende, T. (2002) Fetal development in the pig relation to genetic merit for piglet survival. *J Anim Sci* 80, 1759-1770.

17. Leenhouwers, J., Knol, E., and van der Lende, T. (2002) Differences in late prenatal development as an explanation for genetic differences in piglet survival. *J Anim Sci* 78, 57-62.

18. Voillet, V., SanCristobal, M., Lippi, Y., Martin, P., Iannuccelli, N., Lascor, C., Vignoles, F., Billon, Y., Canario, L., and Liaubet, L. (2014) Muscle transcriptomic investigation of late fetal development identifies candidate genes for piglet maturity. *BMC Genomics* 15, 797.

19. Miller, D. R., D, B., Jackson, R. B., Downie, E. F., and R, R. J. (2010) Metabolic maturity at birth and neonate lamb survival: Association among maternal factors, litter size, lamb birth weight, and plasma metabolic and endocrine factors on survival and behavior. *J Anim Sci* 88, 581-593.

20. Lawn, J. E., Cousens, S., and Zupan, J. (2005) 4 million neonatal deaths: when? where? why? *The Lancet* 365, 891-900.

21. Basso, O. and Wilcox, A. (2010) Mortality risk among preterm babies: immaturity versus underlying pathology. *Epidemiology* 21, 521-527.

22. Foxcroft, G. R., Dixon, W. T., Novak, S., Putman, C. T., Town, S. C., and Vinsky, M. D. (2006) The biological basis for prenatal programming of postnatal performance in pigs. *J Anim Sci* 84, E105-E112.

23. Good, C. A., Kramer, H., and Somogyi, M. (1933) The determination of glycogen. *J Biol Chem* 100, 485-491.

24. Montagne, L., Loisel, F., Le Naou, T., Gondret, F., Gilbert, H., and Le Gall, M. (2014) Difference in short-term responses to a high-fiber diet in pigs divergently selected for residual feed intake. *J Anim Sci* 92, 1512-1523.

25. Perruchot, M. H., Lefaucheur, L., Louveau, I., Mobuchon, L., Palin, M. F., and Farmer, C. (2015) Delayed muscle development in small pig fetuses around birth cannot be rectified by maternal early feed restriction and subsequent overfeeding during gestation. *Animal* 9, 1996-2005.

26. Pfaffl, M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acid Res* 29, e45.

27. Schiaffino, S. and Reggiani, C. (2011) Fiber types in mammalian skeletal muscles. *Physiol Rev* 91, 1447-1531.

28. Lefaucheur, L., Edom, F., Ecolan, P., and Butler-Browne, G. S. (1995) Pattern of muscle fiber type formation in the pig. *Dev Dynam* 203, 27-41.

29. Lefaucheur, L., Hoffman, R., Okamura, C., Gerrard, D., Leger, J. J., Rubinstein, N., and Kelly, A. (1997) Transitory expression of alpha cardiac myosin heavy chain in a subpopulation of secondary generation muscle fibers in the pig. *Dev Dynam* 210, 106-116.

30. Lefaucheur, L., Ecolan, P., Lossec, G., Gabillard, J. C., Butler-Browne, G. S., and Herpin, P. (2001) Influence of early postnatal cold exposure on myofiber maturation in pig skeletal muscle. *J Muscle Res Cell M* 22, 439-452.

31. Vincent, A., Louveau, I., Gondret, F., Trefeu, C., Gilbert, H., and Lefaucheur, L. (2015) Divergent selection for residual feed intake affects the transcriptomic and proteomic profiles of pig skeletal muscle. *J Anim Sci* 93, 2745-2758.

32. Breiman, L. (2001) Random forests. *Mach Learn* 45, 5-32.

33. Lê Cao, K.-A., Boitard, S., and Besse, P. (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12, 253.

34. Lê Cao, K.-A., Rossouw, D., Robert-Granie, C., and Besse, P. (2008) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol* 7, article 35.

35. Witten, D. M., Tibshirani, R., and Hastie, T. (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10, 515-534.

36. Francois, Y., Marie-Etancelin, C., Vignal, A., Viala, D an Davail, S., and Molette, C. (2014) Mule duck 'foie gras' shows different metabolic states according to its quality phenotype by using a proteomic approach. *J Agr Food Chem* 62, 7140-7150.

37. Faure, J., Lefaucheur, L., Bonhomme, N., Ecolan, P., Meteau, K., Coustard, S. M., Kouba, M., Gilbert, H., and Lebret, B. (2013) Consequences of divergent selection for residual feed intake in pigs on muscle energy metabolism and meat quality. *Meat Sci* 93, 37-45.

38. Lebret, B., Le Roy, P., Monin, G., Lefaucher, L., Caritez, J. C., Talmant, A., Elsen, J. M., and Sellier, P. (1999) Influence of the three rn genotypes on chemical composition, enzyme activities and myofiber characteristics of porcine skeletal muscle. *J Anim Sci* 77, 1482-1489.

39. Team, R. C. (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

40. Reverter, A. and Chan, E. (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491-2497.

41. Hudson, N., Reverter, A., Wang, Y., Greenwood, P., and Dalrymple, B. (2009) Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. *PloS ONE* 4, e7249.

42. Pérez-Montarelo, D., Hudson, N., Fernández, A., Ramayo-Caldas, Y., Dalrymple, B., and Reverter, A. (2012) Porcine tissue-specific regulatory networks derived from meta-analysis of the transcriptome. *PloS ONE* 7, e46159.

43. Watson-Haigh, N., Kadarmideen, H., and Reverter, A. (2010) PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics* 26, 411-413.

44. Newman, M. and Girvan, M. (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69, 026113.

45. Reichardt, J. and Bornholdt, S. (2006) Statistical mechanics of community detection. *Phys Rev E* 74, 0161110.

46. Montastier, E., Villa-Vialaneix, N., Caspar-Bauguil, S., Hlavaty, P., Tvrzicka, E., Gonzalez, I., Saris, W., Langin, D., Kunesova, M., and Viguerie, N. (2015) System model network for adipose tissue signatures related to weight changes in response to calorie restriction and subsequent weight maintenance. *PloS Comput Biol* 11, e1004047.

47. Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., and Morris, Q. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acid Res* 1, 214-220.

48. Tabas-Madrid, D., Nogales-Cadenas, R., and Pascual-Montano, A. (2012) GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acid Res* 10, 478-483.

49. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57, 289-300.

50. Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal* Complex Systems, 1695.

51. Laurent, T. and Villa-Vialaneix, N. (2011) Using spatial indexes for labeled network analysis. *Information, Interaction, Intelligence* , 11.

52. Noack, A. (2009) Modularity clustering is force-directed layout. *Phys Rev* 79, 026102.

53. Bastian, M., Heymann, S., and Jacomy, M. (2009) Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* 1, 1:2.

54. Tricarico, D., Selvaggi, M., Passantino, G., De Palo, P., Dario, C., Centoducati, P., Tateo, A., Curci, A., Maqoud, F., Mele, A., Camerino, G. M., Liantonio, A., Imbrici, P., and Zizzo, N. (2016) ATP

sensitive potassium channels in the skeletal muscle function: Involvement of the KCNJ11(kir6.2) gene in the determination of mechanical warner bratzer shear force. *Front Physiol* 7, 167.

55. Chang, K.-C. and Fernandes, K. (1997) Developmental expression and 5' end cDNA cloning of the porcine 2x and 2b myosine heavy chain genes. *DNA Cell Biol* 16, 1429-1437.

56. Picard, B., Lefaucheur, L., Berri, C., and Duclos, J. (2002) Muscle fibre ontogenesis in farm animal species. *Reprod Nutr Dev* 42, 415-431.

57. Lefaucheur, L., Milan, D., Ecolan, P., and Le Callennec, C. (2004) Myosin heavy chain composition of different skeletal muscles in large white and meishan pigs. *J Anim Sci* 82, 1931-1941.

58. Cox, R. D. and Buckingham, M. E. (1992) Actin and myosin genes are transcriptionally regulated during mouse skeletal muscle development. *Dev Biol* 149, 228-234.

59. Herpin, P., Damon, M., and LeDividich, J. (2002) Development of thermoregulation and neonatal survival in pigs. *Livest Prod Sci* 78, 25-45.

60. Rehfeldt, C., Lefaucheur, L., Block, J., Stabenow, B., Pfuhl, R., Otten, W., Metges, C. C., and Kalbe, C. (2012) Limited and excess protein intake of pregnant gilts differently affects body composition and cellularity of skeletal muscle and subcutaneous adipose tissue of newborn and weanling piglets. *Eur J Nutr* 51, 151-165.

61. Petrak, J., Ivanek, R., Toman, O., Cmejla, R., Cmejlova, J., Vyoral, D., Zivny, J., and Vulpe, C. D. (2008) Déjà vu in proteomics. a hit parade of repeatedly identified differentially expressed proteins. *Proteomics* 8, 1744-1749.

62. Wang, P., Bouwman, F. G., and Mariman, E. C. M. (2009) Generally detected proteins in comparative proteomics - a matter of cellular stress response? *Proteomics* 9, 2955-2966.

63. Kornberg, A. and Pricer, W. (1953) Enzymatic esterification of alpha-glycerophosphate by long chain fatty acids. *J Biol Chem* 204, 345-357.

64. Bonneau, M., Mourot, J., Noblet, L., Lefaucheur, L., and Bidanel, J. P. (1990) Tissue development in meishan pigs: Muscle and fat development and metabolism and growth regulation by somatotropic hormone. *Chinese Pig Symp* , 202-213.

65. Lefaucher, L. and Vigneron, P. (1986) Postnatal changes in some histochemical and enzymatic characteristics of three pig muscles. *Meat Sci* 16, 199-216.

66. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19, 1720-1730.

67. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. E. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929-934.

68. Schmidt, M. W., Houseman, A., Ivanov, A. R., and Wolf, D. A. (2007) Comparative proteomic and transcriptomic profiling of the fission yeast schizosaccharomyces pombe. *Mol Syst Biol* 3, 79.

69. Yin, L., Tao, Y., Zhao, K., Shao, J., Li, X., Liu, G., Liu, S., and Zhu, L. (2007) Proteome and transcriptomic analysis of rice mature seed-derived callus differentiation. *Proteomics* 7, 755-768.

70. Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004) Integrated genomic and proteomic analyses of gene expression in mammalian cells. *Mol Cell Proteomics* 3, 960-969.

71. Xun, Z., Sowell, R. A., Kaufman, T. C., and Clemmer, D. E. (2007) Protein expression in a drosophila model of parkinson's disease. *J Proteome Res* 6, 348-357.

72. Ghazalpour, A., Bennett, B., Petyuk, V. A., Orozco, L., Hagopian, R., Mungrue, I. N., Farber, C. R., Sinsheimer, J., Kang, H. M., Furlotte, N., Park, C. C., Wen, P. Z., Brewer, H., Weitz, K., II, D. G. C., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D. J., Smith, R. D., and Lusis, A. J. (2011) Comparative analysis of proteome and transcriptome variation in mouse. *PloS Genet* 7, e1001393.

73. Waters, K. M., Liu, T., Quesenberry, R. D., Willse, A. R., Bandyopadhyay, S., Kathmann, L. E., Weber, T. J., Smith, R. D., Wiley, H. S., and Thrall, B. D. (2012) Network analysis of epidermal growth factor signaling using integrated genomic, proteomic and phosphorylation data. *PloS ONE* 7, e34515.

74. Wu, Z., Puigserver, P., Andersson, U., Zhang, C., Adelmant, G., Mootha, V., Troy, A., Cinti, S., Lowell, B., Scarpulla, R., and Spiegelman, B. (1999) Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* 98, 155-124.

75. Jäger, S., Handschin, C., St-Pierre, J., and Spiegelman, B. M. (2007) Amp-activated protein kinase (ampk) action in skeletal muscle via direct phosphorylation pf pgc-1alpha. *P Natl Acad Sci USA* 104, 12017-12022.

76. Oberkofler, H., Esterbauer, H., Linnemayr, V., Strosberg, A., Krempler, F., and Patsch, W. (2002) Peroxisome proliferator-activated receptor (ppar) alpha coactivator-1 recruitment regulates ppar subtype specificity. *J Biol Chem* 277, 16750-16757.

77. Chan, M. C. and Arany, Z. (2014) The many roles of PGC-1alpha in muscle–recent developments. *Metabolism* 63, 441-451.

78. Lin, J., Wu, H., Tarr, P., Zhang, C., Wu, Z., Boss, O., Michael, L., Puigserver, P., Isotani, E., Olson, E., Lowell, B., Bassel-Duby, R., and Spiegelman, B. (2002) Transcriptional co-activator pgc-1alpha drives the formation of slow-twitch muscle fibres. *Nature* 418, 797-801.

79. Zhao, R., Yang, X., Xu, Q., Wei, X., Xia, D., and Chen, J. (2004) Expression of GHR and PGC-1alpha in association with changes of myhc isoform types in longissimus muscle of erhualian and large white pigs (sus scrofa) during postnatal growth. *Anim Sci* 79, 203-211.

80. Wende, A. R., Schaeffer, P. J., Parker, G. J., Zechner, C., Han, D. H., Chen, M. M., Hancock, C. R., Lehman, J. J., Huss, J M nd McClain, D. A., Holloszy, J. O., and Kelly, D. P. (2007) A role for the transcriptional coactivator PGC-1alpha in muscle refueling. *J Biol Chem* 282, 36642-51.

81. Tcherepanova, I., Puigserver, P., Norris, J. D., Spiegelman, B. M., and McDonnell, D. P. (2000) Modulation of estrogen receptor-alpha transcriptional activity by the coactivator PGC-1. *J Biol Chem* 275, 16302-16308.

82. Hamilton, D. J., Minze, L. J., Kumar, T., Cao, T. N., Lyon, C. J., Geiger, P. C., Hsueh, W. A., and Gupte, A. A. (2016) Estrogen receptor alpha activation enhances mitochondrial function and systemic metabolism in high-fat-fed ovariectomized mice. *Physiol Rep* 4, e12913.

83. Heine, P. A., Taylor, J. A., Iwamoto, G. A., Lubahn, D. B., and Cooke, P. S. (2000) Increased adipose tissue in male and female estrogen receptor-alpha knockout mice. *Proc Natl Acad Sci USA* 97, 12729-12734.

84. Nilsson, M., Dahlman, I., Rydén, M., Nordström, E. A., Gustafsson, J. A., Amer, P., and Dahlman-Wight, K. (2007) Oestrogen receptor alpha gene expression levels are reduced in obese compared to normal weight females. *Int J Obesity* 31, 900-907.

85. Ribas, V., Audrey Nguyen, M. T., Henstridge, D. C., Nguyen, A. K., Beaven, S. W., Watt, M. J., and Hevener, A. L. (2010) Impared oxidative metabolism and inflammation are associated with insulin resistance in er alpha-deficient mice. *Am J Physiol-Endoc M* 298, 304-319.

86. Lefaucheur, L. (2010) A second look into fibre typing–relation to meat quality. *Meat Sci* 84, 257-270.

87. Brocks, L., Klont, R., Buist, W., de Greef, K., Tieman, M., and Engel, B. (2000) The effects of selection of pigs on growth rate vs leanness on histochemical characteristics of different muscles. *J Anim Sci* 78, 1247-1254.

Figure Legends

Figure 1: Overview of the network data integration analysis.

(A) Proteome networks were first inferred using the PCIT (partial correlation and information theory) algorithm for each developmental time point: 90 days of gestation and 110 days of gestation. PCIT, which belongs to the family of weighted network algorithms, is based on the combination of the concept of partial correlation coefficient and the information theory to identify meaningful associations. **(B)** Network clustering was then performed using a spin-glass algorithm. A permutation test was performed to assess the significance of clustering. **(C)** Sub-networks (clusters) were analyzed. (i) GeneMANIA was used to assess the relevance of the proposed sub-networks and GO enrichment analysis was performed to analyze the biological functions of the sub-networks. (ii) Centrality (or betweenness) of nodes was analyzed using a permutation test. (iii) Using spatial statistics tools, correlations between sub-networks and biological phenotypes of interest were analyzed.

Figure 2: Co-variation of muscle glycogen and expression profiles (mRNA and proteins) of myosin heavy chain (MyHC) genes (adult fast (IIa + IIb + IIx) and embryonic MyHC).

Lowess curves and confidence intervals are in blue and grey, respectively. Box-plots of mRNA or protein expression and glycogen content are also shown. **(A)** mRNA embryonic MyHC. **(B)** Protein embryonic MyHC. **(C)** mRNA adult Fast (IIa + IIb + IIx) MyHC. **(D)** Protein adult Fast (IIa + IIb + IIx) MyHC.

Figure 3: Global analysis of the 113 identified proteins.

Venn diagram of the statistical methods chosen to select spots for identification. 13 proteins (in blue) were differentially expressed according to the interaction between developmental time points and fetal genotypes. 58 proteins (in red) were differentially expressed in an additive manner between development time points and fetal genotypes. 62 proteins (in green) were selected using sparse multivariate methods (both sCCA and sPLS). 55 proteins (in yellow) were selected using discriminant analyses (both RF and sPLS-DA). See the Materials and Methods section for further information.

Figure 4: Level plot of the 113 identified spots.

All spots were ranged according to the biological function. Muscle development functions and components,

e.g. system development, actin skeleton, muscle filament sliding, cytoskeleton and mRNA metabolic process, were mostly over-expressed at 90 dg. Energy metabolism functions and components, *e.g.* glucose and gluconeogenesis metabolism, oxidation-reduction activity and mitochondrion, were mostly over-expressed at the end of gestation (110 dg). All the biological functions and components are not included in this figure.

Figure 5: Decomposition of the proteome networks.

(A) Each sub-network represents a cluster in the d90-proteome network. **(B)** Each sub-network represents a cluster in the d110-proteome network. The width of the edge is proportional to the number of edges between the two corresponding sub-networks.

Figure 6: Three sub-networks obtained at 90 days of gestation.

A PCIT model was used to infer a co-expression network. Nodes were clustered using a spin-glass model. The figure shows three (out of five) sub-networks (1, 2 and 4) obtained by clustering. The color of the nodes indicates betweenness centrality. The red nodes have the highest betweenness. The size of the nodes indicates degree. The color of the edges indicates the correlation sign: red for positive correlations and blue for negative ones.

Figure 7: Box plots of protein level expression.

PSMC5 with two isoforms, GPD1 with four isoforms, ATP5A1 and CKMT2.

Figure 8: Representative western blot gel (A) and box plot analysis (B) of CKMT2 expression in longissimus muscle (n = 64 fetuses).

M: molecular weight marker. Different letters between groups within fetal stages denote significant differences ($P < 0.05$). ***: significant difference between fetal stages ($P < 0.001$).

Figure 9: Enzyme activities of lactate dehydrogenase (LDH), citrate synthase (CS) and beta-hydroxy-Acyl-CoA dehydrogenase (HAD) in longissimus muscle of LWLW and MSMS fetuses at 90 and 110 dg.

Activities are expressed as μmol of substrate degraded per minute per gram fresh muscle. Different letters

denote significant differences ($P < 0.05$).

Figure 10: Five sub-networks obtained at 110 days of gestation.

A PCIT model was used to infer a co-expression network. Nodes were clustered using a spin-glass model. This figure shows five (out of six) sub-networks (2, 3, 4, 5 and 6) obtained by clustering. The color of the nodes indicates betweenness centrality. The red nodes have the highest betweenness. The size of the nodes indicates degree. The color of the edges indicates the correlation sign: red for positive correlations and blue for negative ones.

Figure 11: Correlation between transcriptome and proteome in both extreme fetal genotypes (LW and MS).

Scatter plot of correlations between mRNA and protein expression levels in MS (x-axis) and LW (y-axis). Each dot represents the correlation in MS and LW of a pair of gene-protein. Red dots show a significant positive correlation in LW and MS, the green dots show a significant positive correlation only in MS, and blue dots show a significant positive correlation only in LW. It is important to note that more than 81 points were represented because some proteins had several isoforms.

Figure 12: Bibliographic network obtained with the d110-sub-network 2. This network is composed of 30 nodes. Ten nodes out of the 24 composing sub-network 2 were found to be possibly transcriptionally regulated. Box plots of PPARGC1A, KCNJ11, MYOG and ESR1 mRNA expression from [18] are also shown.

Tables

	90 days of gestation				110 days of gestation					P-values[2]		
	LW	MSLW	LWMS	MS	LW	MSLW	LWMS	MS	RSD[1]	A	G	AxG
MyHC Protein level, % total MyHC												
Embryonic	16.20^{a}	12.71^{b}	10.31^{c}	9.25^{c}	4.10^{d}	2.62^{e}	2.10^{e}	1.53^{e}	0.94	<.001	<.001	<.001
Perinatal	65.81^{a}	68.32^{b}	71.51^{c}	71.56^{c}	47.52^{d}	45.38^{d}	43.88^{d}	39.31^{e}	2.57	<.001	0.120	<.001
Fast (IIa + IIx + IIb)	9.06^{a}	10.07^{a}	9.99^{a}	9.90^{a}	32.28^{b}	37.10^{c}	39.91^{c}	43.77^{d}	2.82	<.001	<.001	0.003
I + α-cardiac	8.93^{a}	8.89^{a}	8.20^{a}	9.30^{a}	16.14^{b}	14.92^{b}	14.10^{b}	15.41^{b}	1.17	<.001	0.013	0.448
MyHC RNA level												
Embryonic	4.66^{a}	3.58^{b}	2.23^{c}	2.21^{c}	1.18^{d}	0.39^{d}	0.36^{d}	0.19^{d}	0.54	<.001	<.001	0.008
Perinatal	1.49^{a}	1.73^{a}	1.66^{a}	1.76^{a}	0.82^{b}	0.62^{b}	0.68^{b}	0.59^{b}	0.21	<.001	0.987	0.010
Fast (IIa + IIx + IIb)	0.11^{a}	0.17^{a}	016^{a}	0.18^{a}	0.84^{b}	1.05^{c}	1.21^{c}	1.80^{d}	0.17	<.001	<.001	<.001
IIa	0.12^{a}	0.16^{a}	0.14^{a}	0.17^{a}	0.76^{b}	1.09^{c}	1.27^{c}	1.87^{d}	0.22	<.001	<.001	<.001
IIx	0.12^{a}	0.18^{a}	0.18^{a}	0.19^{a}	0.92^{b}	0.95^{b}	1.14^{b}	1.73^{c}	0.27	<.001	0.002	0.006
I	0.53^{a}	0.60^{a}	0.47^{a}	0.61^{a}	1.02^{b}	0.99^{b}	0.98^{b}	1.60^{c}	0.20	<.001	<.001	<.001
α-cardiac	0.09^{a}	0.13^{a}	0.14^{a}	0.14^{a}	1.85^{b}	2.72^{b}	3.48^{b}	6.46^{c}	1.46	<.001	<.001	<.001
Glycogen, % wet muscle	3.38^{a}	4.03^{a}	4.68^{b}	4.99^{b}	9.55^{c}	10.52^{d}	12.36^{e}	12.11^{e}	0.58	<.001	<.001	0.033

Table 1. Biological phenotype characteristics of the 64 fetuses according to the eight experimental conditions.
The 64 fetuses used for this analysis were the same fetuses as those used in the proteome and transcriptome analyses. We considered eight conditions according to two developmental time points (d90 and d110) and four fetal genotypes (MS, LW, MSLW and LWMS). [1]Residual Standard Deviation (or Standard error of estimate). [2]A, Age; G, Genotype; AxG, Age x Genotype interaction. ^{a,b,c,d,e} Means with different letters are significantly different ($P < 0.05$).

47

90 days of gestation					
Community	**Nb. of nodes**	**Nb. of edges**	**Important node**[1]	**Phenotype link**[2]	**Gene Ontology**[3]
1	18	26	FETUB	Glycogen & Emb. MyHC	glycolysis, myofibril, mitochondrion
2	19	55	PSMC5 (Iso 2)*	Glycogen & Fast MyHC	gluconeogenesis, actin filaments, mitochondrion
3	21	47	PDIA3 (Iso 1)	-	respiratory electron transport chain, cell redox homeostasis, mitochondrion
4	20	49	TNNT3 (Iso 2)*	Emb. MyHC	gluconeogenesis, muscle filament sliding and development, cell cycle checkpoint
5	16	33	ANXA2 (Iso 2)	-	creatine metabolic process, sarcolemma

Table 2. Network characteristics at 90 days of gestation.

Five sub-networks (communities) were obtained at 90 days of gestation. [1]Nodes with high betweenness. [2]The phenotype correlations, with MyHC embryonic, adult fast and muscle glycogen content, were analyzed using spatial statistics tools. [3]Additional File 10 shows all significantly enriched GO terms for each sub-network. *Using a permutation test, betweenness significantly high ($P < 0.05$).

48

110 days of gestation					
Community	**Nb. of nodes**	**Nb. of edges**	**Important nodes[1]**	**Phenotype correlation[2]**	**Gene Ontology[3]**
1	5	5	TUFM, DLD	-	tricarboxylic acid cycle, mitochondrion
2	24	65	ATP5A1*	Fast & Emb. MyHC	tricarboxylic acid cycle, glucose metabolic process, muscle filament sliding
3	19	46	GSN (Iso 1)	Glycogen, Fast & Emb. MyHC	gluconeogenesis, glycolysis, muscle filament sliding
4	13	19	PSMC5 (Iso 1)	Glycogen & Emb. MyHC	mRNA metabolic process, proteasome complex
5	15	39	LDB3 (Iso 2)	Glycogen & Fast MyHC	muscle development, lipid metabolism
6	10	22	PSMC5 (Iso 2)	Glycogen	gluconeogenesis, ATP catabolic/anabolic process, mitochondrion

Table 3. Network characteristics at 110 days of gestation.

Six sub-networks (communities) were obtained at 110 days of gestation. [1]Nodes with a high betweenness. [2]The phenotype correlations, with MyHC embryonic, adult fast and muscle glycogen content, were analyzed using spatial statistics tools. [3]Additional File 12 shows all significantly enriched GO terms for each sub-network. *Using permutation test, betweenness significantly high ($P < 0.05$).

49

Figure

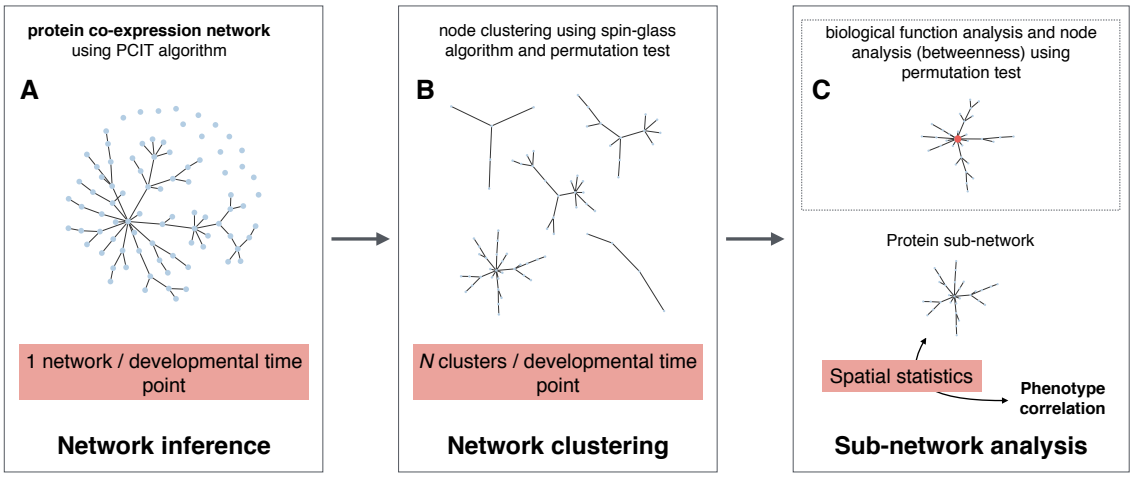

Figure 1

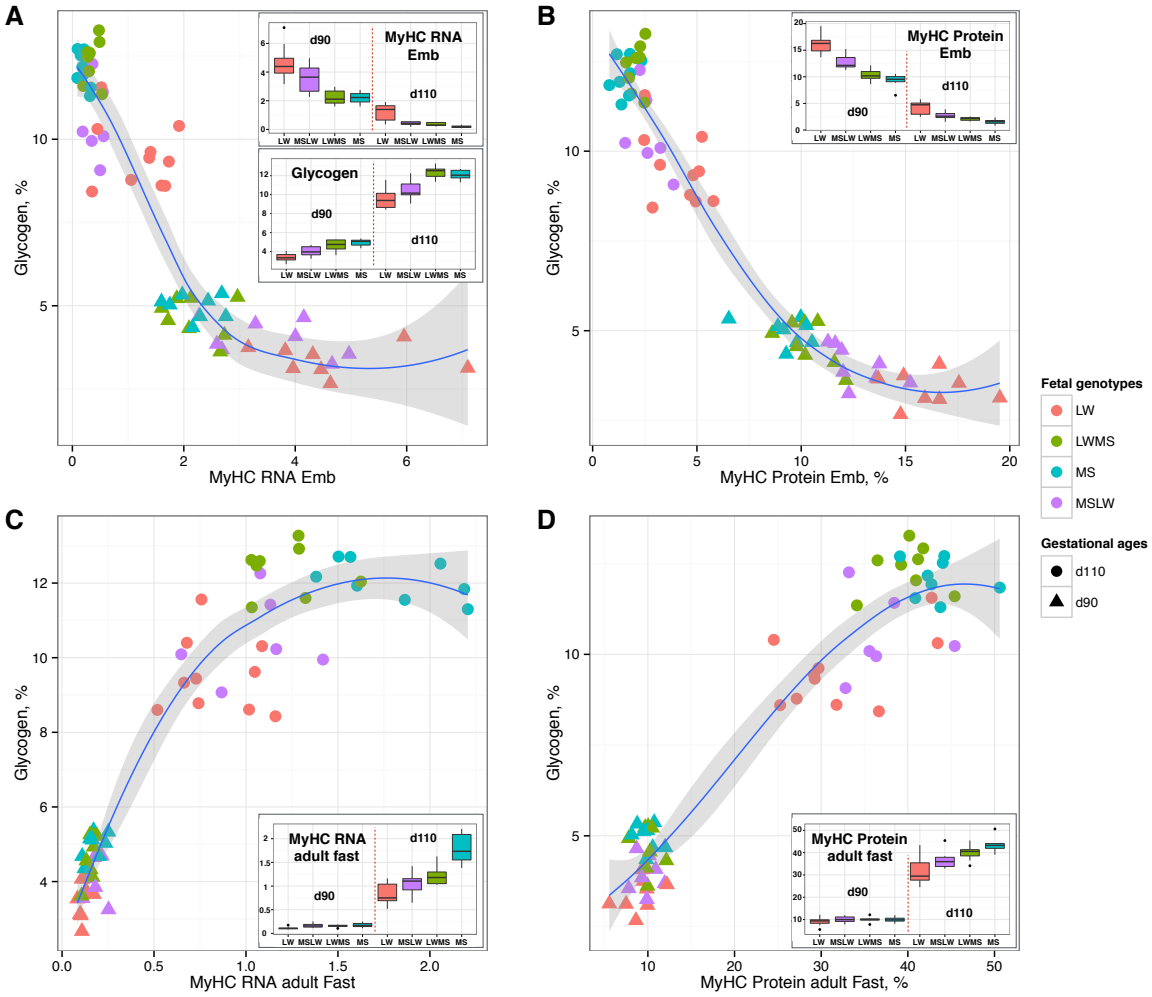

Figure 2

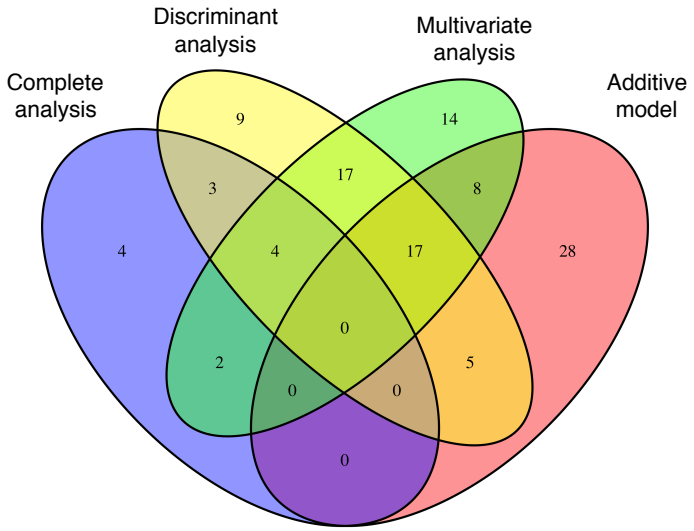

Figure 3

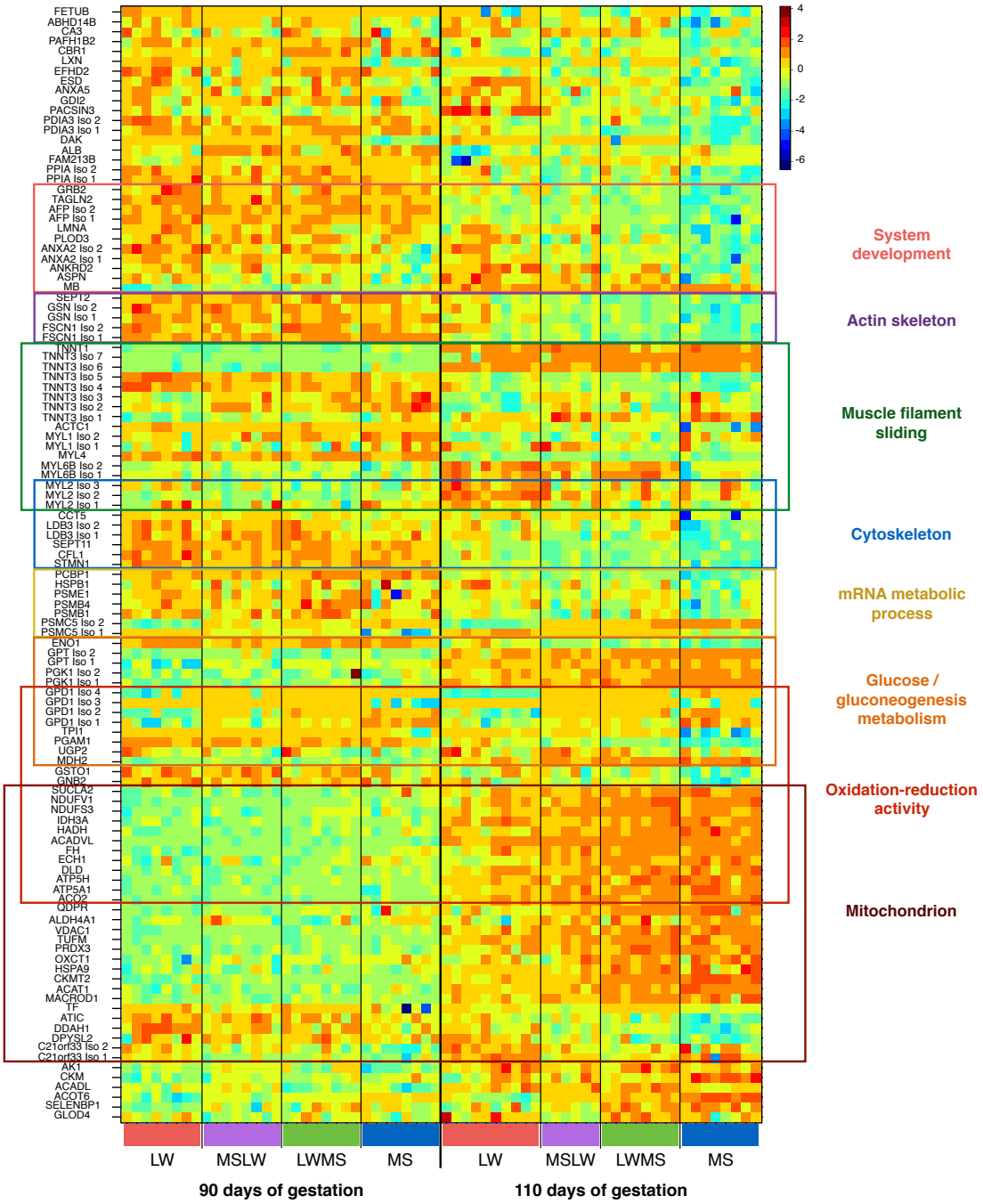

Figure 4

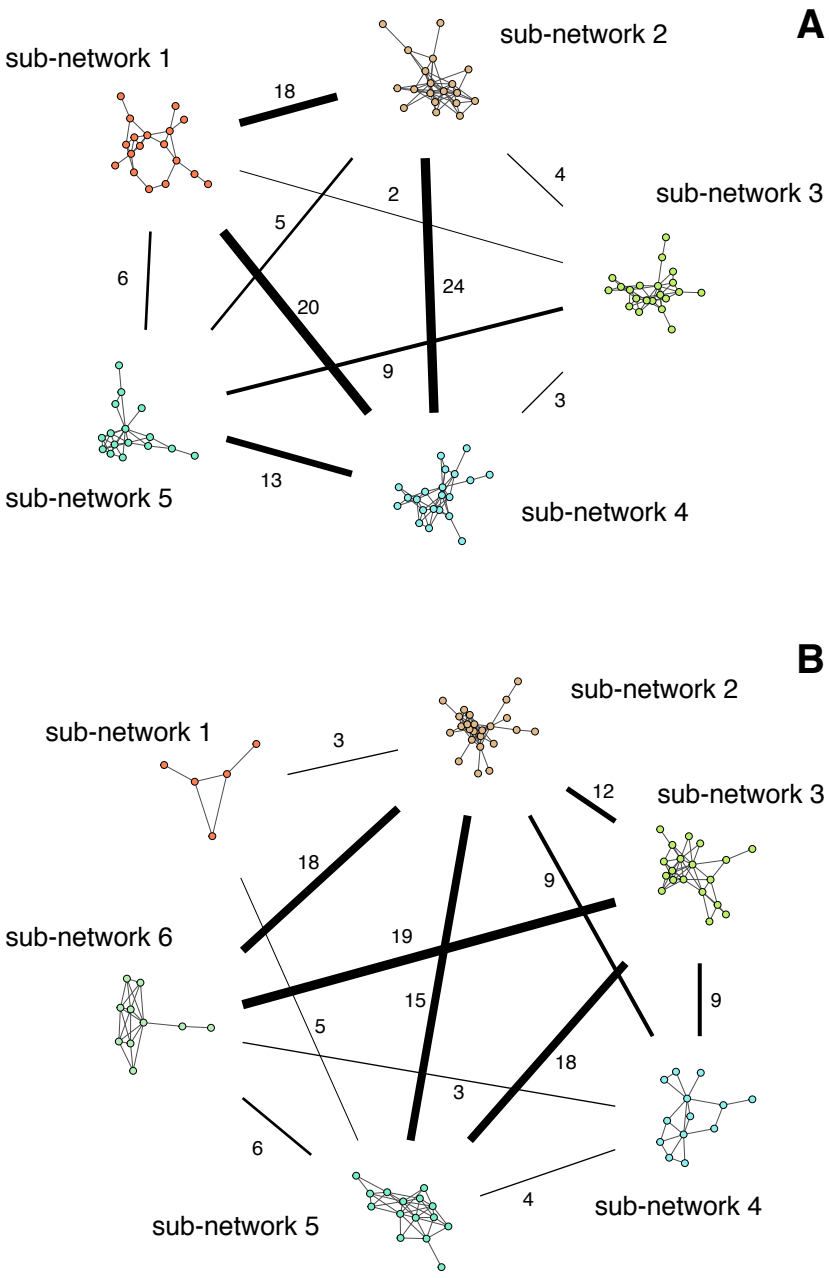

Figure 5

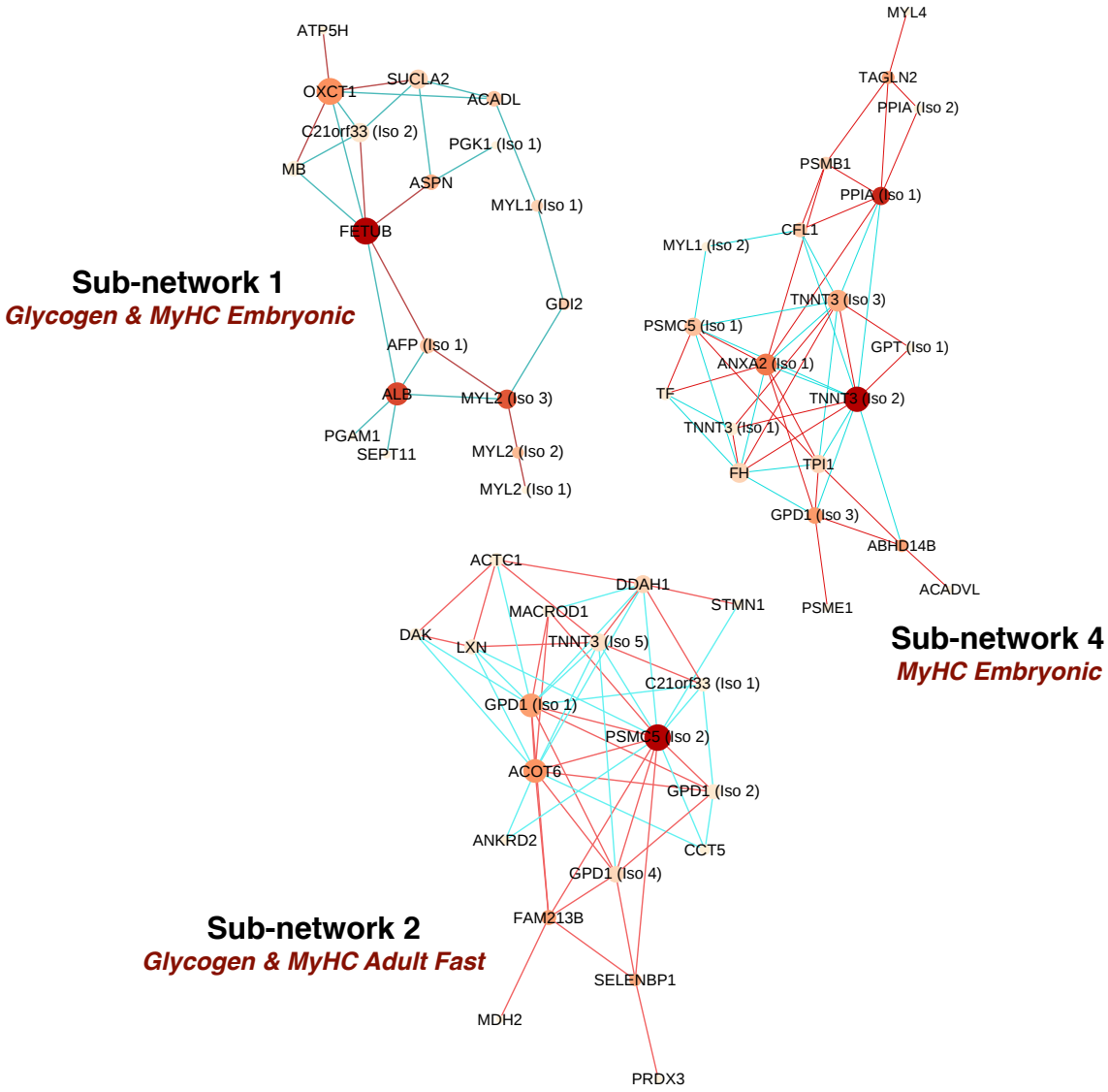

Figure 6

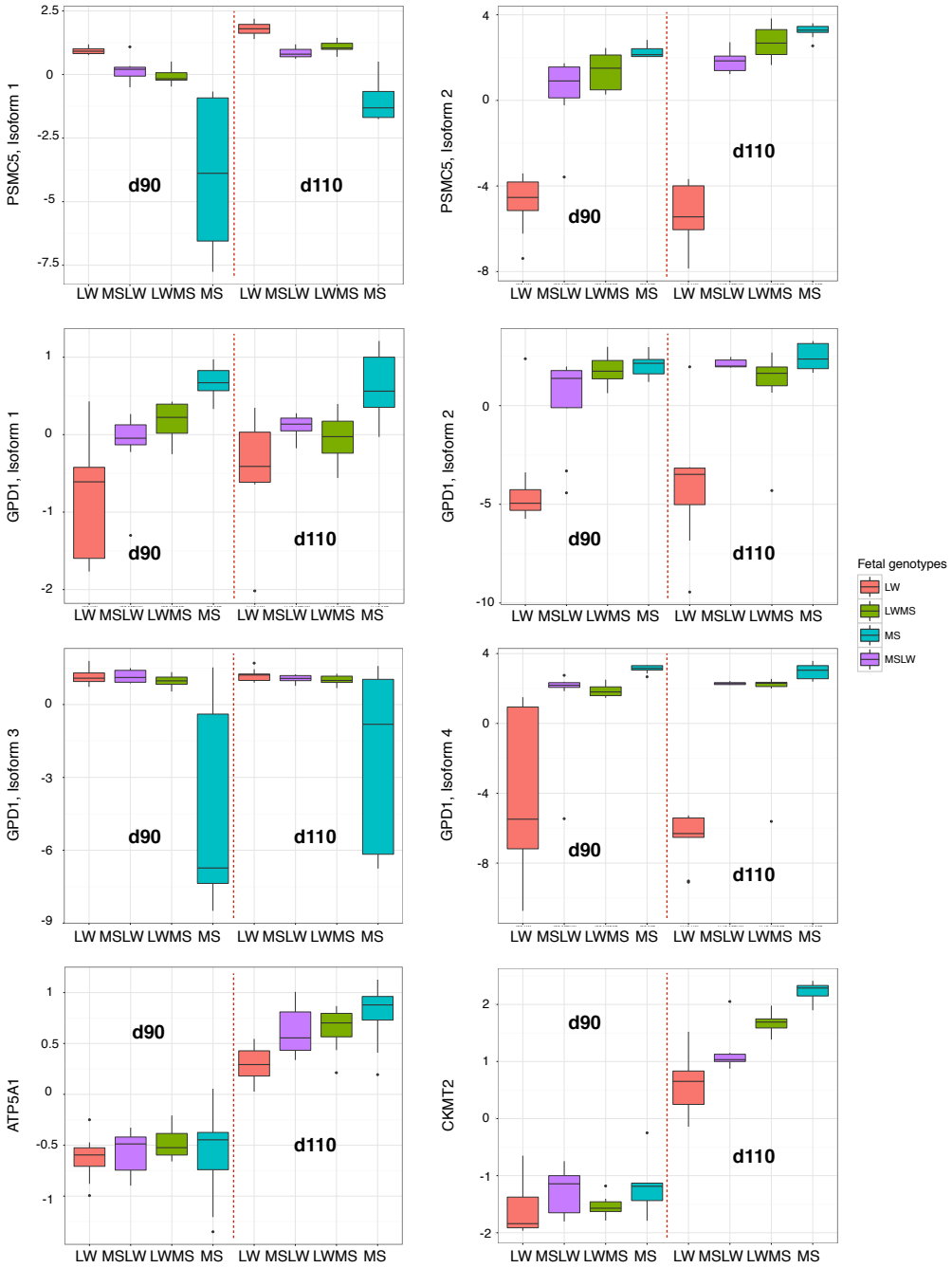

Figure 7

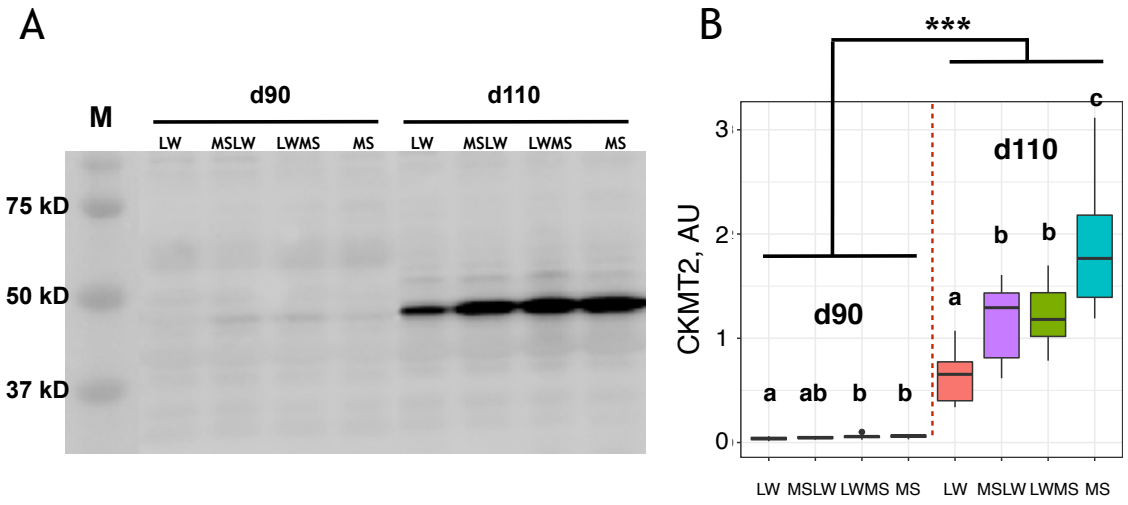

Figure 8

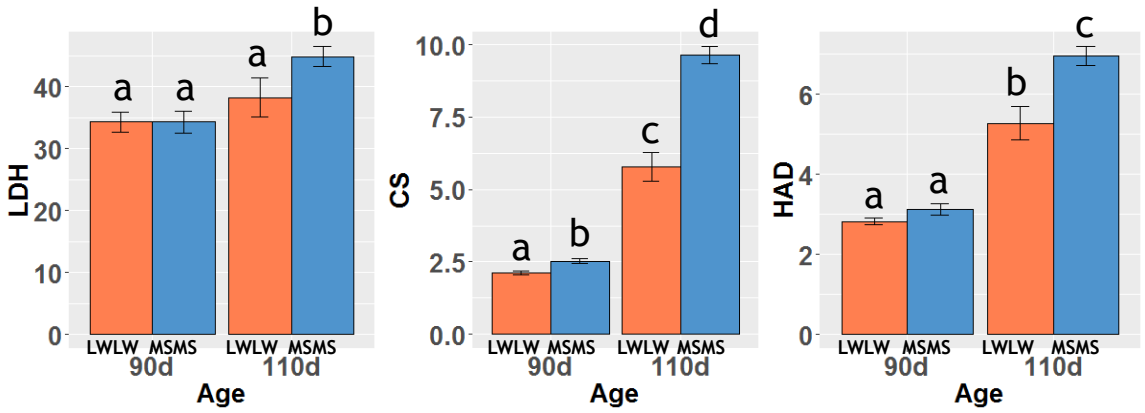

Figure 9

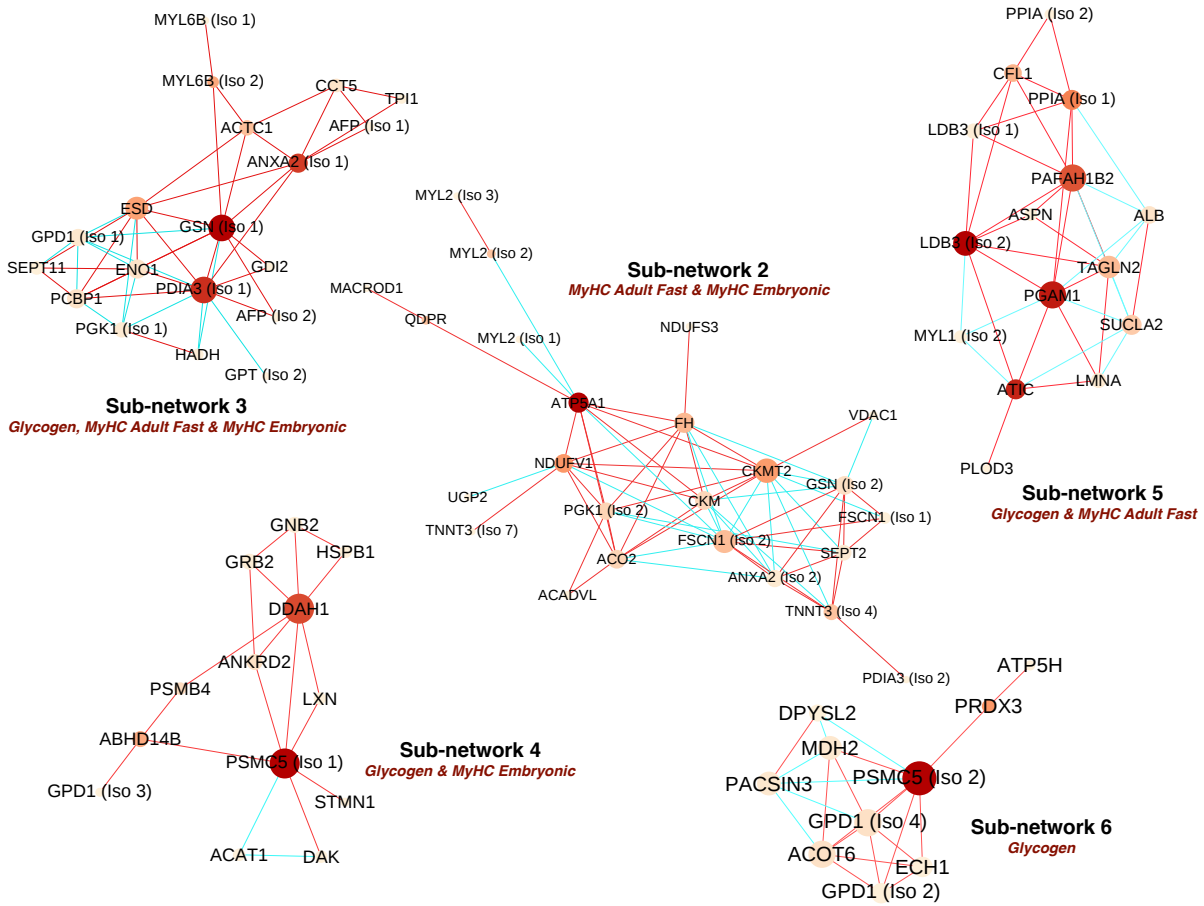
MYL6B (Iso 1)
MYL6B (Iso 2)
CCT5
TPI1
ACTC1
AFP (Iso 1)
ANXA2 (Iso 1)
ESD
GSN (Iso 1)
GPD1 (Iso 1)
SEPT11
ENO1
GDI2
PCBP1
PDIA3 (Iso 1)
AFP (Iso 2)
PGK1 (Iso 1)
HADH
GPT (Iso 2)
Sub-network 3
Glycogen, MyHC Adult Fast & MyHC Embryonic
MYL2 (Iso 3)
MYL2 (Iso 2)
MACROD1
QDPR
MYL2 (Iso 1)
NDUFS3
Sub-network 2
MyHC Adult Fast & MyHC Embryonic
ATP5A1
FH
VDAC1
NDUFV1
CKMT2
GSN (Iso 2)
UGP2
CKM
PGK1 (Iso 2)
FSCN1 (Iso 1)
TNNT3 (Iso 7)
FSCN1 (Iso 2)
SEPT2
ACO2
ANXA2 (Iso 2)
ACADVL
TNNT3 (Iso 4)
PPIA (Iso 2)
CFL1
PPIA (Iso 1)
LDB3 (Iso 1)
PAFAH1B2
ASPN
ALB
LDB3 (Iso 2)
TAGLN2
PGAM1
SUCLA2
MYL1 (Iso 2)
LMNA
ATIC
PLOD3
Sub-network 5
Glycogen & MyHC Adult Fast
GNB2
HSPB1
GRB2
DDAH1
ANKRD2
PSMB4
LXN
ABHD14B
PSMC5 (Iso 1)
GPD1 (Iso 3)
STMN1
ACAT1
DAK
Sub-network 4
Glycogen & MyHC Embryonic
PDIA3 (Iso 2)
ATP5H
PRDX3
DPYSL2
MDH2
PACSIN3
PSMC5 (Iso 2)
GPD1 (Iso 4)
ACOT6
ECH1
GPD1 (Iso 2)
Sub-network 6
Glycogen

Figure 10

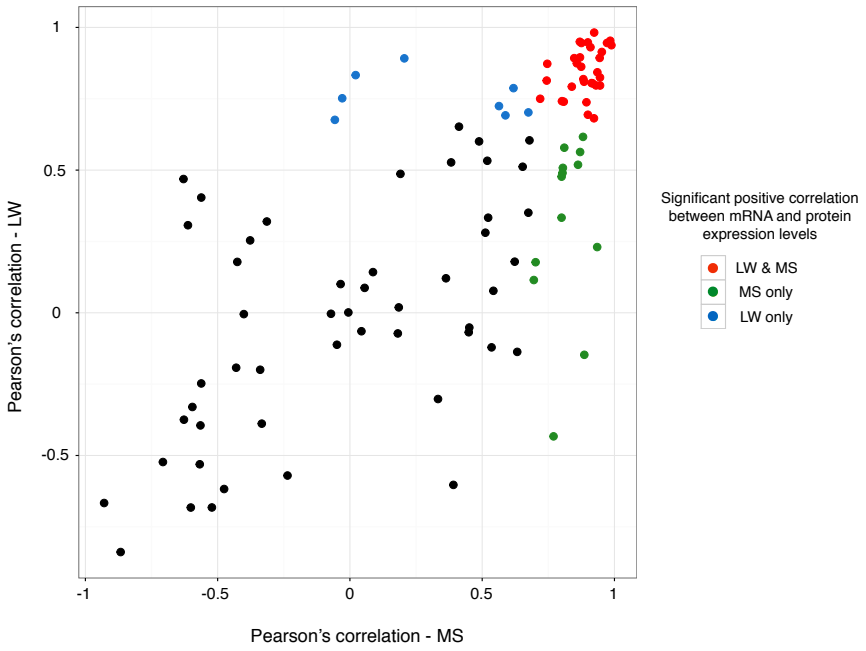

Figure 11

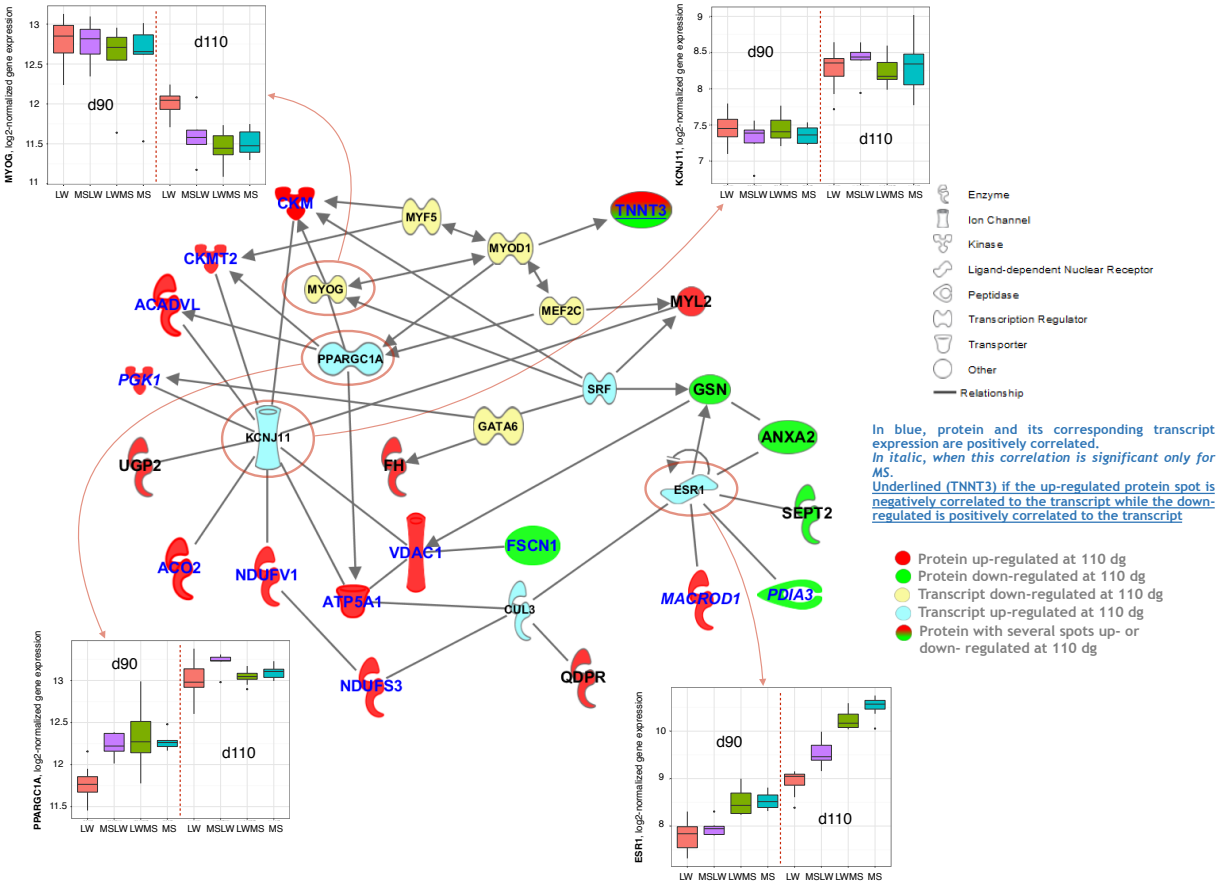
MYOG, log2-normalized gene expression
d90
d110
LW MSLW LWMS MS LW MSLW LWMS MS
KONJ11, log2-normalized gene expression
CKM
MYF5
TNNT3
CKMT2
MYOD1
MYOG
ACADVL
MEF2C
MYL2
PPARGC1A
PGK1
SRF
GSN
KCNJ11
GATA6
ANXA2
UGP2
FH
ESR1
SEPT2
ACO2
NDUFV1
VDAC1
FSCN1
ATP5A1
CUL3
MACROD1
PDIA3
NDUFS3
QDPR
PPARGC1A, log2-normalized gene expression
ESR1, log2-normalized gene expression
Enzyme
Ion Channel
Kinase
Ligand-dependent Nuclear Receptor
Peptidase
Transcription Regulator
Transporter
Other
Relationship
In blue, protein and its corresponding transcript expression are positively correlated.
In italic, when this correlation is significant only for MS.
Underlined (TNNT3) if the up-regulated protein spot is negatively correlated to the transcript while the down-regulated is positively correlated to the transcript
Protein up-regulated at 110 dg
Protein down-regulated at 110 dg
Transcript down-regulated at 110 dg
Transcript up-regulated at 110 dg
Protein with several spots up- or down- regulated at 110 dg

Figure 12