Fat and Muscle DB: integrating 'omics' data from adipose tissue and muscle

Jérémy Tournayre, Isabelle Cassar-Malek, Matthieu Matthieu.Reichstadt@inrae.Fr Reichstadt, Brigitte B. Picard, Nicolas Kaspric, Muriel Bonnet

Farm animal proteomics 2014

Proceedings of the 5th Management Committee Meeting and
4th Meeting of Working Groups 1, 2 & 3 of COST Action FA 1002
Milano, Italy – 17-18 November 2014

edited by:
André de Almeida
Fabrizio Ceciliani
David Eckersall
Ingrid Miller
Jenny Renaut
Paola Roncada
Romana Turk

cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

Fat&MuscleDB: integrating 'omics' data from adipose tissue and muscle

Jérémy Tournayre[1,2], Isabelle Cassar-Malek[1,2], Matthieu Reichstadt[1,2], Brigitte Picard[1,2], Nicolas Kaspric[1,2] and Muriel Bonnet[1,2*]*

[1]INRA, UMR1213 Herbivores, 63122 Saint-Genès-Champanelle, France

[2]Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, 63000, Clermont-Ferrand, France; jeremy.tournayre@clermont.inra.fr; muriel.bonnet@clermont.inra.fr

Introduction

Increasing the amount of meat while preserving a high quality is an economic challenge for the beef industry. From a biological point of view, this aim depends on the relative amounts of muscle and adipose tissues that determines the lean-to-fat ratio. Understanding how rearing practices affect the lean-to-fat ratio has implications for the quality and value of the carcass and the meat from cattle. It had thus motivated vast amounts of transcriptomic and proteomic studies to identify which genes and proteins control tissue growth and physiology. We hypothesize that integration of available published data through a computational strategy will foster discoveries of the genes or proteins that are systematically related to stages of the development of adipose tissue and muscle, as candidate markers of potential tissue growth. Moreover, the assembly of contextualized knowledge will improve our understanding of the interactions between adipose tissue and muscle. Indeed, most of the 'omics' studies questioned the growth of either muscles or adipose tissues in cattle depending on intrinsic (genotype, sex) or environmental (nutrition) factors (Picard *et al.* 2011). They were never merged to question the developmental and functional links between muscle and adipose tissue. These links are suggested by the successive waves of growth of muscle and adipose tissue as well as the wide plasticity of body composition depending on age or genotype in cattle (Bonnet *et al.*, 2010). Currently, some functional links are rarely reported in monogastrics thanks to *in vitro* studies (Romacho *et al.*, 2014).

To gather available data, we designed the Fat&MuscleDB, an online database of transcriptomic and proteomic data from skeletal muscle and adipose tissue. Because muscular and adipose masses grow by hyperplasia during foetal or early post-natal life in cattle as well as by hypertrophy, we integrated data related to these mechanisms both *in vivo* and *in vitro*, from ruminants but also from models species or cell lines experiments. We defined criteria relative to the experiments and the stages of adipose tissue and muscle growth in order to record and classify data. Our final objective is to aggregate all references that share the same criterion related to a stage of adipose tissue and muscle growth to help biologists to find candidate proteins to be biomarkers of the lean-to-fat ratio.

Methods

A database of transcripts and proteins (with their abundance values) dealing with the growth of adipose and muscle tissue has been built from Pubmed and Gene Expression Omnibus (GEO) on the National Center for Biotechnology Information (NCBI) and the Web of Knowledge. To select contextualized articles or dataset, keywords were defined by experts in fat and muscle growth and submitted via an automatic procedure. Each retrieved publication (with its supplementary data) was analysed manually to check for results and table(s) reporting transcripts or proteins differentially abundant during muscle and/or fat growth. The extraction of tables from a PDF was simplified by the use of a tool that has been implemented. We used 'pdftotext' command to convert PDF to text and then to a tabbed file. Transcriptomics results of GEO were analysed to determine transcripts differentially expressed or not during muscle and/or fat growth. Curators had chosen, among a GEO Series (GSE), GEO Samples (GSM) to submit to analyse in the context of adipose tissue and muscle growths. Transcriptomics results were then recovered and processed using a tool developed in R software with libraries GEOquery (Davis and Meltzer, 2007) and Anapuce (http://cran.rproject.org/web/packages/anapuce). Anapuce was chosen for its specific functions ('DiffAnalysis' and 'DiffAnalysis.unpaired') for the differential analysis between two groups of microarray data. Data were normalized when necessary through 'normalisation' function, spots for a same transcript and technical replicates were averaged and the variance mixture was calculated with 'DiffAnalysis' or 'DiffAnalysis.unpaired' functions for paired data or unpaired data, respectively. We chose a p-value of 0.15 for the variance because of the stringency of this test that strongly reduced the number of regulated transcripts. Each of the data analysed by these methods was verified by curators and annotated as 'Analysed by Fat&MuscleDB'. In contrast, data recorded as they were published were annotated as 'Not analysed by Fat&MuscleDB'. Data were recorded with information given by the authors regarding the experimental protocols in order to allow investigators to conduct researches and aggregations in Fat&MuscleDB.

In order to record, classify and aggregate data, we defined 73 criteria on experiments and stages of adipose tissue and muscle growth. The aggregation of data implies to have a unique identifier for a gene or a protein. We have chosen to convert all names or identifiers in UniProt accession from UniProt which is a protein database regularly updated and curated by experts. To do this, each name or identifiers was sent on the search engine of UniProt or NCBI in order to get the UniProt accession for the species investigated, otherwise in any species for not losing information. By this way, the aggregation process produced a table that merged results from all experiments sharing a same criterion, and all transcripts or proteins were referred by a UniProt accession. Each accession was counted once by reference in order to identify the redundancy of identification of a gene or a protein.

Results and discussion

Three lists of keywords have been created to seek references from PubMed, GEO and Web of Knowledge in order to establish a reference database. One list described cellular features of adipose and muscle tissues (hypermuscularity, high adiposity, myoblast, adipocyte...), another characterizes species and cell lines (bovine, human, 3T3-L1, C2C12) and the latest relates methods (transcriptomics, proteomics). Association of keywords between these lists has generated approximately 27,000 combinations that were automatically launched on public databases. Moreover, in order to refine the search, negative keywords have been added to each association. Following this search, we found 15,552 publications and 2,312 GEO datasets. Right now, 2,366 publications and 928 GEO datasets does not contain information that interests us contrary to 185 publications and 139 GEO datasets must be processed to be inserted into the database. Among those to be treated, we recovered 164 side-by-side comparisons from 26 publications and 18 GEO datasets in order to identify genes or proteins with an stable, increased or decreased abundance. We continue to feed Fat&MuscleDB, there are still over 250 references on all species to treat and over 12,000 references to analyse.

We report the example of the integration of data related to the criterion 'Muscular hypertrophy from genetic origin' (Figure 1). By clicking on the 'Aggregation' button, investigators have to select this criterion among others and to select other settings to submit the integration. Currently this integration uses data from six references analysed by twelve comparisons. Thus, UniProt accession lists are built up of transcripts and proteins according to their abundance. After the integration of the data by the aggregation process we identified 13 proteins and 357 transcripts with a decreased abundance, 3 proteins and 286 transcripts with an increased abundance and 17,088 transcripts with a stable abundance that are related to 'Muscular hypertrophy from genetic origin'. This list will be submitted for data mining using informatics resources such as ProteINSIDE that we have developed (Kaspric *et al.*, 2014), to identify central or relevant genes or proteins related to muscular hypertrophy. Aggregation can be achieved for the 72 other criteria that we have defined to depict all the stages of adipose tissue and muscle growth.

The time devoted to analysing references is substantial, so references dealing with ruminants were analysed in priority. However, because proteins and transcripts underlying fat and muscle growth can be the same in several species, we will complete our database with data from human and cells lines. Fat&MuscleDB provides an innovative and interactive resource that connects beef industry interests with the new biological discoveries in 'omics' research. This computational strategy will catalyze the assembly of knowledge about muscle and adipose tissue growth, and foster the discovery of candidate biomarkers of the lean-to-fat-ratio. This would also minimize unnecessary redundancy in research efforts.

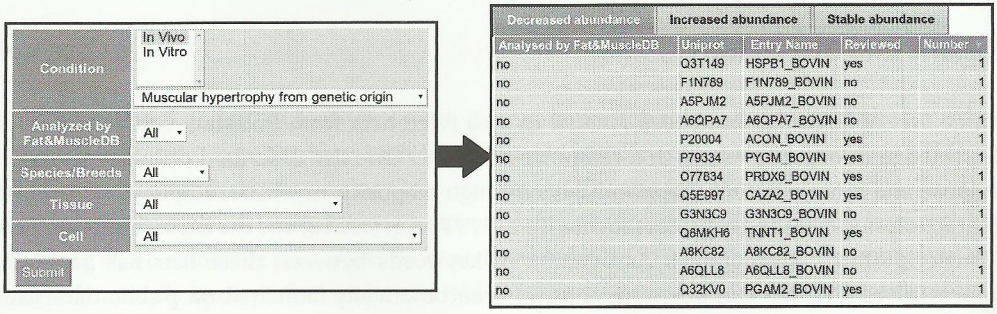

Figure 1. Integration of 'omics' data relative to 'Muscular hypertrophy from genetic origin' on Fat&MuscleDB. (Left) The form that allows selecting references to aggregate according to different settings: in vitro or in vivo experiment, if the data were analysed by us or recovered according to author interpretation, breeds, species, tissues and lines cell. (Right) Extract of the list of proteins whose abundance decreases obtained by aggregating data for 'Muscular hypertrophy from genetic origin'. Each line corresponds to a protein with multiple indications: if it was recovered by us ('Yes') or recovered according to author interpretation ('No'), its UniProt accession, its Entry name, if it has been reviewed by UniProt consortium or not, and the number of times it is aggregated (a maximum of one by reference). Here, all proteins were recorded by only one reference.

Acknowledgements

This study was supported by the regional council of Auvergne in France through the regional information system Lifegrid.

References

Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B., 2010. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. Animal 4: 1093-1109.

Davis, S. and Meltzer, P.S., 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 23: 1846-1847.

Kaspric, 2014. Proteomics data from ruminants easily investigated using ProteINSIDE. In: De Almeida, A., *et al.* (ed.), 2014. Farm animal proteomics 2014 – Proceedings of the 5th Management Committee Meeting and 4th Meeting of Working Groups 1,2 & 3 of COST Action FA 1002, Wageningen Academic Publishers, Wageningen, the Netherlands, pp. 283-287.

Picard, B., Cassar-Malek, I., Guillemin, N. and Bonnet, M., 2011. Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: M. Moo-Young (ed.), Comprehensive Biotechnology (Second Edition). Academic Press, Burlington, pp. 395-405.

Romacho, T., Elsen, M., Rohrborn, D. and Eckel, J., 2014. Adipose tissue and its role in organ crosstalk. Acta Physiol (Oxf) 210: 733-753.

Proteomics is the large-scale study of the proteome, i.e. a set of proteins being expressed in a certain fluid, tissue, organ or organism. The value of this advanced technology is being recognised in farm animal and veterinary sciences from 'farm to fork'. The potential of proteomics is unequivocal in holding a significant promise in applications such as vaccine and drug development, physiology, toxicology, animal product quality and food safety. Proteomics has been growing steadily during the last 3-4 years and, as time goes by, proteomics-based studies are more and more common, not just to scientists but to the general public as well, unravelling the full potential of this innovative technology.

This book reflects the will of a group of multi-disciplinary scientists that merge innovation with excellence of research and to whom the dissemination of knowledge and discovery through cooperation is a key point. It is of interest to scientists at the early stages of their careers as well as to researchers well established in the field and to whom proteomics may be the necessary next step towards more in-depth research activities. By providing a collection of diverse scientific interests, *Farm Animal Proteomics 2014* demonstrates the vitality of the area and the importance it holds to animal and food research, to science, industry, government agencies, the consumer and ultimately the society as a whole.

ISBN 978-90-8686-262-7

ESF provides the COST Office through an EC contract. COST is supported by the EU RTD Framework Programme

Wageningen Academic Publishers