

Intégration de données publiques et analyses protéomiques pour révéler les mécanismes et biomarqueurs du dépôt de lipides dans la carcasse ou le muscle bovin

Jeanne Bazile

► To cite this version:

Jeanne Bazile. Intégration de données publiques et analyses protéomiques pour révéler les mécanismes et biomarqueurs du dépôt de lipides dans la carcasse ou le muscle bovin. Agronomie. Université Clermont Auvergne [2017-2020], 2019. Français. NNT : 2019CLFAC074 . tel-03155659

HAL Id: tel-03155659 https://theses.hal.science/tel-03155659

Submitted on 2 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.







ECOLE DOCTORALE DES SCIENCES DE

LA VIE ET DE LA SANTE – AGRONOMIE - ENVIRONNEMENT

THESE

présentée pour obtenir le grade de **DOCTEUR D'UNIVERSITE**

Spécialité : Génétique, Physiologie, Pathologie, Nutrition, Microbiologie, Santé, Innovation

Soutenue publiquement le **18 décembre 2019**, devant la commission d'examen.

Par Jeanne BAZILE

Intégration de données publiques et analyses protéomiques pour révéler les mécanismes et biomarqueurs du dépôt de lipides dans la carcasse ou

le muscle bovin

Membres du Jury :

| Pr Jean-Marc LOBACCARO (Professeur, GReD – Clermont-Ferrand) | Président |
|---|---------------------|
| Dr Cécile BERRI (Directeur de recherche, INRA – Nouzilly) | Rapporteur |
| Dr Sébastien DEJEAN (Ingénieur de Recherche, Université Paul Sabatier – Toulouse) | Rapporteur |
| Dr Isabelle HUE (Chargé de recherche, INRA – Rennes) | Examinateur |
| Dr Etienne PAUX (Directeur de recherche, INRA – Clermont-Ferrand) | Examinateur |
| Dr Muriel BONNET (Directeur de recherche, INRA – Clermont-Ferrand) | Directrice de thèse |

Préparée dans l'Unité Mixte de Recherche sur les Herbivores, Equipe Biomarqueurs INRA Centre Auvergne Rhône-Alpes

"Around here, however, we don't look backwards for very long. We keep moving forward, opening up new doors and doing new things, because we're curious... and curiosity keeps leading us down new paths." *Walt Disney*

Remerciements

Je voudrais tout d'abord remercier les personnes qui ont accepté d'évaluer ce travail : Jean-Marc LOBACCARO, qui a accepté d'être président de ce jury, Cécile BERRI et Sébastien DEJEAN pour avoir accepté d'être rapporteurs de ce manuscrit ainsi que Étienne PAUX et Isabelle HUE pour avoir aimablement accepté d'être évaluateurs.

Merci également aux personnes qui ont bien voulu lors de ces 3 ans aiguiller au mieux mes travaux de thèse lors de mes comités de thèse. Que ce soit **Julien MAUPETIT, Bérénice BATUT, Isabelle HUE, Frédéric CHOULET, Denis LALOE, Florence JAFFREZIC, Pierre NOZIERE, Maguy EUGENE** (pour ses conseils en amont et en aval de ces comités) et **Brigitte PICARD** : merci à vous tous pour votre aide précieuse et votre expertise.

Un grand merci ensuite à mon encadrante de thèse, **Muriel BONNET**. Merci de m'avoir donné ma chance. Tu as eu confiance en moi quand cela me faisait défaut et su me remettre sur le droit chemin quand je commençais à me perdre. Tu as toujours su te rendre disponible malgré des emplois du temps parfois compliqués, et particulièrement en fin de thèse où tu t'es donnée à 200 %. Je reste impressionnée par ta capacité à traiter plusieurs sujets à la fois en ayant une vision très précise et très globale à la fois. Tu as toujours été sincère avec moi et j'ai été ravie de travailler avec toi. Tu m'as poussé à me dépasser et j'ai découvert des ressources dont je n'avais même pas idée.

Un grand merci ensuite à **Laurence BERNARD**. Je n'aurais pas pu choisir meilleure tutrice. Tu as toujours été là en soutien et tes conseils ont été plus que précieux. Tu as su être à l'écoute et me rassurer quand cela était nécessaire et m'a même aiguillé pour faire disparaître mes maux de tête (ce qui n'était pas une mince affaire). Pour cela je te suis éternellement reconnaissante. Merci également pour ta relecture attentive de ma thèse. Et merci pour les conversations dans le bus et pour tes petits passages dans notre bureau pour voir comment nous allions !

Un grand merci également à **Florence JAFFREZIC** et **Denis LALOE**, ainsi qu'à leur équipe dans laquelle je me suis senti bien dès mon arrivé et même si je ne restais pas très longtemps. C'était un véritable plaisir de travailler avec vous 2 et je vous remercie d'avoir accepté cette collaboration et de votre implication et conseils dans ce projet. Merci également à **Patrice DEHAIS** et **Christophe KLOPP** pour leur aide sur la mise à jour, leurs explications et les propositions faites pour le traitement de données. Je remercie ensuite l'équipe Biomarqueurs pour son superbe accueil. C'était un plaisir de vous retrouver tous les matins. Merci de m'avoir acceptée, aidée, soutenue, et ce tous les jours : **Carole DELAVAUD**, **Agnès THOMAS**, **Denys DURAND**, **Brigitte PICARD**, **Céline BOBY** (vers l'infini et l'au-delà de la bio-informatique), **Christine LEROUX**, **Dominique GRUFFAT** pour tes réponses aux questions les plus inattendues et pour ta relecture attentive, **Isabelle CASSAR-MALEK** (merci pour le séjour australien !), **José PIRES** (oui cousin, j'ai réussi à « poublier » sans ton nom et sans ta « poèle »), **Marie-Pierre ELLIES**, **Yannick FAULCONNIER** (pour ta joie de vivre et ton implication dans l'organisation de l'ISNH qui était au top ! C'était un plaisir de vous donner un coup de main !), **Arnaud DELAVAUD**, **David « Crousti » CHADEYRON**, **Martine TOURRET** et ton fameux fauteuil et ton rire inimitable, **Nicole DUNOYER**, **Quentin DELAHAYE** et tes succulents sushis !, **Sébastien BES** et **Sylvie BARDOU-VALETTE** pour ta bonne humeur.

Petit mot spécial à **Mylène DELOSIERE**, co-bureau temporaire, toujours là pour donner un coup de main, ton avis ou juste faire du soutien moral !

Je remercie également mes co-bureaux : **Sabrina BOUDON** pour les discussions sportives et les souvenirs australiens, **Hélène FOUGERE** et surtout **Julien SOULAT**/Captain Juju/Bichon. Merci en particulier à Julien pour les conversations sur tout et sur rien (pas mal sur rien d'ailleurs), le soutien et conseils quelle que soit la question, mon diplôme de traduction en Julien, mais surtout ta patience pour nous supporter (à moins que ce soit nous qui te supportions ??). Je te souhaite un avenir vachement bien !

Merci également à **Sylvain** « Sysy » **EMERY** 007, pour ta patience également, tes histoires de chaussettes, ta bonne humeur, et ton aide pour me remonter le moral !

Je remercie également les autres non-titulaires de Biomarqueurs : **Pierre-Alexis BILLA** notre sonneur de soupe officiel, **Daoud OUNAISSI**, **Alexandre CONANEC** et surtout **les** « **Piou-Piou** » **Thomas LÉONARDON et Benjamin RIBEIRO**. Cela m'amène aux nontitulaires des autres équipes de l'UMRH. Merci à vous tous pour les moments passés ensemble. Des remerciements notamment pour **Églantine THIERY** et nos discussions matinales du bus, **Madeline KOCZURA** pour ses histoires, **Pauline DENIS** et **Lucille REY** (ou « les bébés Peraq »)

Des remerciements spéciaux pour 2 personnes qui me sont particulièrement chères : **Louise FORTEAU** et **Solveig MENDOWSKI**, mes partenaires du trio de choc. Nous finirons comme nous avons commencé : ensemble. Un grand merci pour ces 3 ans de soutien, de rires, et d'organisations diverses ! Cela a été un plaisir de partager ces 3 ans avec vous et je vous souhaite beaucoup de bonheur à toutes les 2. Merci Solveig de m'avoir permis de refaire du cheval avec Un Bord (3). D'autres soutiens précieux sont ceux outre-Atlantique : **Hélène FOUGERE** pour les petits messages canadiens, les rires partagés dans le bureau, les conseils (avec **Bénédict YANIBADA**) et ta gentillesse ! Merci également à **Morgane BONNOT-GILLARD** (ou plutôt Morgane de « Titouan et Morgane »), pour les petits messages américains pour conseiller, prendre des nouvelles ou juste encourager. Et enfin à la 2^e canadienne, **Camille** LOISEAU et sa petite famille, un gros bisou.

Merci à l'équipe Peraq également, ainsi que les personnes avec qui je mange le midi, notamment **Pascale BERAUD** (« tu l'as vu ? »), **Vincent LARGEAU** et ses histoires racontées comme personne avec **Nicolas GERARD**, **Cyril LABONNE**, **Didier BANY**, **Émilie TIXIER** et **Luciano BARRETO-MENDES** pour vos gâteaux de folies, **Yoan GAUDRON**, **Mélissa VAURIS**, **Jordann DOMAGALSKI** : merci de m'avoir appris la coinche (bon, je suis très débutante, mais j'ai retenu qu'il fallait être bruyant !), pour vos rires et vos histoires !

De manière générale, merci aux gens de l'Arche, de PI : mention spéciale pour **Matthieu REICHSTADT** pour son aide malgré des demandes plus ou moins possibles, sa réactivité et la découverte du Toucan, ainsi que pour le « co »-encadrement de **Pierre-Baptiste COUNENC.** Merci aussi à GAME, et notamment les fées de GAME **Annie ROYET** et **Nathalie GOCEL** qui m'ont simplifié la vie tout au long de mon doctorat.

À tous ceux que j'ai pu oublier, ainsi que les acteurs de l'ombre (agent de cuisine et de nettoyage et chauffeurs de bus), un grand merci pour votre aide de tous les jours, aussi infime soit-elle.

Je souhaite remercier également les 2 premières personnes à m'avoir fait confiance et à m'avoir formé au travail de scientifique: **Grzegorz KUDLA** et **Olivier DA INES**.

Je remercie mes amis de Clermont : **Medhi, Thibault, Kevin et Corentin** (bref, le gang des monocles), la communauté des bioinfos de Clermont-Ferrand, **Laura** et **Maryline** que je suis toujours ravie de retrouver. À ma family bisontine, **Poussin et Poussinette, Mémé, Debby-Justine et Pupuce** : vous me manquez !! J'espère que nous trouverons l'occasion de nous revoir vite !

Un grand merci à ma famille d'adoption : **Marie-José** et **Vitor DE OLIVEIRA** : Un grand merci pour votre aide et pour les bons moments passés ensemble. J'ai de la chance d'être tombé sur de beaux-parents aussi adorables! Merci également à **Amélie** (et **Izzy**), **Tante Rose**, **Adeline**, **Tonton Jean-Pierre**, **Tante Marie-Laure**, **Marie-Lou**....Merci à tous pour m'avoir accepté dans votre famille.

Merci à ma famille pour leur soutien indéfectible. À mes 2 sœurs chéries, **Marion et Chloé**, qui ont fait du soutien moral où qu'elles soient, de leur amour et qui m'ont appris à être battante et que tout était possible ! Merci à tous pour vos cartes postales qui m'ont accompagnées sur toute la fin d'écriture de ce manuscrit. Merci à **Guillaume** aussi de supporter Chloé...euh, je veux dire pour me supporter au côté de Chloé. Je suis vraiment heureuse pour vous et vous m'avez apporté un petit peu plus de bonheur encore. Je vous souhaite tout le bonheur possible ! À mes deux parents, je ne sais vous dire à quel point je vous aime. Vous m'avez toujours portée, supportée, poussée, soutenue quels que soient mes choix et aidée dans mes décisions. À ma Mamoun, pêcheuse de homard émérite et mon Papou, chef de tout et « Tu es mon père » (sortie du dernier Star Wars le jour de ma thèse oblige :-P). Merci pour vos appels, les aides diverses, les avis, les remontages de moral express ou juste les nouvelles. Je ne sais vous dire à quel point je vous suis reconnaissante pour votre amour et soutien continuels. Je vous dédie cette thèse.

Enfin à mon chéri **Romain**, mon coach, mon ami, mon partenaire, mon sauveteur, mon doudou. Tu as été mon phare dans les différentes tempêtes, un conseiller de tous les instants. Sans toi, je n'aurais pas tenté cette aventure et j'aurais manqué de grandes (et petites) joies, la réalisation de certains rêves et la fierté de me dépasser aussi bien sportivement qu'intellectuellement. Nous avons déjà parcouru un bout de chemin ensemble, littéralement ou non, avec des hauts et des bas, mais toujours ensemble. J'espère que cela va durer encore longtemps.

Résumé

L'équilibre entre les masses des tissus musculaire (TM) et adipeux (TA) conditionne le poids de carcasse et son rendement (rapport TM et TA), mais aussi les qualités sensorielle et nutritionnelle des viandes. Comprendre comment contrôler le rapport des masses de muscle relativement à celle des tissus adipeux reste un enjeu majeur pour les filières de production bovine. Les méthodes omiques ont abondamment été utilisées pour comprendre les mécanismes explicatifs de la variabilité du croît ou de la masse des tissus adipeux et musculaire chez le bovin. Cependant, à partir de cette masse de données, il n'est pas toujours aisé d'extraire ou de générer une information biologique synthétique. L'objectif de ma thèse était donc d'agréger et de ré-analyser des données publiques pour proposer des gènes ou protéines marqueurs de l'adiposité corporelle ou musculaire, compléter les connaissances disponibles ou identifier des manques de données. Pour ce faire, des techniques in silico et expérimentale ont été combinées.

La majorité des données disponibles dans les bases publiques concernaient le muscle et étaient des données transcriptomiques, rarement protéomiques. Les résultats de 5 publications comparant le protéome musculaire de bovins divergents pour la teneur en lipides intramusculaires ont été agrégés pour identifier 50 protéines avec des abondances divergentes, dont 9 trouvées dans au moins 2 publications. Ces données ayant été obtenues chez des races précoces, nous avons analysé le protéome d'une race plus tardive. Le muscle *Longissimus thoracis* de bovins, Rouges des Prés divergents par les adiposités musculaire et corporelle ont été analysés par les techniques de shotgun et de 2DE. Sur les 47 protéines liées significativement aux dépôts adipeux dans le muscle ou la carcasse, 21 étaient communes avec les résultats publiés, et 26 n'avaient jamais été identifiées. En particulier, l'abondance d'APOBEC2 était fortement corrélée à l'adiposité de la carcasse et du muscle.

Parmi les données issues de puces à ADN accessibles dans les bases de données publiques, 84 et 12 jeux de données relatifs, respectivement à la croissance musculaire et adipeuse, ont été sélectionnés. En raison de manques dans les métadonnées, seulement 33 (32 « TM », 1 « TA ») d'entre eux ont été rendus comparables par la mise à jour des noms de gène sur le génome bovin actuel (UCD1.2 ; collaboration avec Sigenae). Les données des 32 jeux de données « muscle » ont été catégorisées selon l'âge, la race, le sexe ou l'alimentation. Par catégorie, les données ont été regroupées et analysées par combinaison de valeurs p selon la méthode inverse normale (collaboration avec l'UMR Gabi). Pour la catégorie âge, facteur de variation majeur de la teneur en lipides intramusculaires, nous avons identifié 238 gènes (sur les 715 communs à 10 jeux de données) différentiellement exprimés dans le longissimus dorsi de bovins de 5 races. Parmi ces 238 gènes, 97 étaient différentiels dans au moins 2 des jeux de données analysés individuellement et la méta-analyse. Ils confirmaient par exemple la dynamique de régulation des métabolismes glycolytiques et oxydatif en fonction de l'âge des bovins. 17 gènes ont été identifiés comme différentiel selon l'âge exclusivement dans la métaanalyse. Parmi les gènes identifiés certains sont liés au métabolisme des lipides (APOE, LDLR, MXRA8) et d'autres pourraient induire (YBX1) ou réprimer (MAPK14, YWAH, ERBB2), la différenciation de cellules progénitrices musculaires vers le lignage adipeux.

L'intégration de données publiques, notamment par la méta-analyse, a donc augmenté les connaissances sur les mécanismes biologiques et biomarqueurs contribuant au rapport TA/TM. Les relations entre les abondances des molécules identifiées et des critères d'adiposité restent à quantifier dans une perspective de validation des biomarqueurs.

Abstract

Producing meat animals with adequate muscular and adipose masses (i.e. lean-to-fat ratio) is an economic challenge for the beef industry. The lean-to-fat ratio influence weight and yield of carcasses as well as the sensorial and nutritional quality of the meat. Omics methods have been widely used to understand mechanisms underlying the variability of the adipose and muscle tissue growth in the bovine. However, it is not always easy to extract or generate a synthetic biological information from this volume of data. **The objective of my thesis was to aggregate and analyse public data to propose genes or proteins related to the lean-to-fat, and to identify data to be completed by experiment.** To achive this goal, experimental and *in silico* methods were combined.

Majority of available data in public databases were muscle transcriptomic data, and very rarely proteomic data. Data from 5 publications comparing muscle proteome of bovine breeds divergent in their intramuscular lipid content were aggregated and allowed the identification of 50 differently abundant proteins. Of these, 9 were concordant in at least 2 publications. As those data were obtained only in late-maturing breeds, we analysed proteome of "Rouges des Prés" cows which deposit fat in a later stage of development. *Longissimus thoracis* muscle of bovine "Rouges des prés", that diverges in their muscular and corporal adiposity were analysed by shotgun and 2DE techniques. Of the 47 proteins significatively associated to adipose depots in muscle or carcass, 21 were common to published data and 26 had never been identified before. Particularly, APOBEC2 abundance was strongly correlated with both carcass and muscle adiposity.

Among the microarray data available in the public databases, 84 and 12 datasets relative to muscular and adipose growth were selected, respectively. Because of missing metadata, only 33 (32 "MT" and 1 "AT") were used and their identifiers updated on the current bovine genome (UCD1.2; collaboration with Sigenae). Data of 32 "muscle" datasets were categorized according to the age, breed, sex or nutrition. Data were regrouped by categories and analysed by p-value combination according to the inverse normal method (collaboration with Gabi UMR). For the age category, a major factor influencing intramuscular lipid content, we identified 238 genes differentially expressed between two ages in longissimus dorsi of bovine of 5 different breeds. Among these 238 genes, 97 were identified in at least 2 datasets analysed individually and in the meta-analysis. The meta-analysis confirmed the dynamic regulation of glycolytic and oxidative metabolisms depending on bovine age. 17 genes were exclusively identified genes, some are linked to lipid metabolism (APOE, LDLR, MXRA8) and other may induce (YBX1) or repress (MAPK14, YWAH, ERBB2) the differentiation of muscle progenitor cells towards the adipose lineage.

Integration of public data, in particular by the meta-analysis, provided a global view of biological mechanisms and biomarkers (genes or proteins) of the lean-to-fat ratio that were the most frequently identified in several breeds. The relationships between the abundances of the identified molecules and adiposity criteria remain to be quantified in a perspective of biomarker validation (prior to animal slaughter/ without the need to slaughter the animal).

Integration of public data, in particular by the meta-analysis, allowing to aquire a global view of biological mechanisms and biomarkers (genes or proteins) of the AT to MT ratio. This approach will allow to check and evaluate carcass and muscle AT percentage from gene or protein abundance prior to animal slaughter/ without the need to slaughter the animal.

LISTE DES TRAVAUX

Articles en préparation ou soumis dans des journaux à comité de lecture

Bazile, J., Picard, B., Chambon, C., Valais, A. et Bonnet, M. (2019). Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gelbased and gel-free proteomic analyses. Meat science, 156(May), 146-155. doi:10.1016/j.meatsci.2019.05.018

Ceciliani, F., Lecchi, C., Bazile, J. et Bonnet, M. (2018). Proteomics Research in the Adipose Tissue. Dans Proteomics in Domestic Animals: from Farm to Systems Biology (p. 233-254). Cham : Springer International Publishing. doi:10.1007/978-3-319-69682-9_12

Bazile, J., Jaffrezic, F., Laloe, D., Dehais, P., Klopp, C., Bonnet M.,. Molecular signatures of muscle growth and composition deciphered by the meta-analysis of age-related public transcriptomics data. En préparation pour physiological genomics

Communications scientifiques

Posters

- Bazile, J., Jaffrezic, F., Laloe, D. et Bonnet, M. (2018). Meta-analysis of public transcriptomics data to understand and phenotype bovine body composition. R. Baumont, M. Silberberg et I. Cassar-Malek (dir.), Communication présentée au 10th International Symposium on the Nutrition of Herbivores, Clermont-Ferrand. doi:10.13140/RG.2.2.13920.74244
 - ISNH (10th International Symposium on the Nutrition of Herbivores): 2 au 6 septembre 2018 présentation du poster, *Clermont-Ferrand, France*
 - Journée de l'école doctorale SVSAE : 14 et 15 juin 2018 : Présentation du poster en 1 minute, *Clermont-Ferrand, France*

Communications orales

Bazile, J., Bonnet, M., Aggregation of genomic public data to identify biomarkers of bovine body composition, 26 octobre 2017, présentation de 10 minutes devant des collègues du CSRIO, *Brisbane, Australie*

- Bazile, J., Bonnet, Mining of public data and databases, Meta-analysis of transcritpomics data to identify molecular signatures of body composition, 2-6 septembre 2018, ISNH, *Clermont-Ferrand, France*
- Bazile, J., Bonnet, présentation de mes travaux à la journée annuelle des doctorants et post-doctorants de l'UMRH, présentation de 5 minutes (2017) et de 10minutes (2018 et 2019)

Table des matières

| | Avan | t-propo | S | 1 |
|---------|--------|-------------|--|----------|
| | Chap | itre 1 : I | Productions et qualités des carcasses et viandes bovine | 5 |
| | 1 F | roduction | et qualité des produits des bovins | 6 |
| | 1.1 | Productio | on et consommation de la viande bovine en France | 6 |
| | 1.2 | Outils d'é | valuation de la qualité des carcasses à l'abattoir | 8 |
| | 1.3 | Les outils | d'évaluation de la qualité de la viande | 11 |
| | 1.4 | Les facte | urs de variations influençant les adiposités corporelle et musculaire des bovins | 13 |
| | 1 | .4.1 L'â | ge | 14 |
| | | 1.4.1.1 | Croissance du tissu musculaire | 15 |
| | | 1.4.1.2 | Croissance du tissu adipeux | 16 |
| | | 1.4.1.3 | es compétitions lors de la croissance des tissus adipeux et musculaire | 18 |
| | 1 | .4.2 La | ace | 20 |
| | 1 | .4.3 Le s | sexe | 24 |
| | 1 | .4.4 L'a | imentation | 25 |
| | 2 C | Detention e | t disponibilité des données transcriptomiques et protéomiques chez le bovin | 27 28 |
| | 2 | .1.1 Pro | téomique basée sur les gels | 28 |
| | 2 | .1.2 Pro | téomique sans gel | 28 |
| | 2.2 | Les méth | odes pour l'obtention de données transcriptomiques | 29 |
| | 2 | .2.1 Puc | e à ADN | 29 |
| | 2 | .2.2 Séq | uençage ARN | 30 |
| | 3 C | onnées pr | otéomiques appliquées à la compréhension de l'adiposité corporelle et tissulai | re |
| des bov | vins | | | 33 |
| | 4 C | onnées tra | nscriptomiques pour la compréhension de l'adiposité corporelle ou tissulaire d | les |
| bovins | | | | 35 |
| | Chap | itre 3 : | | 41 |
| | La Met | a-analyse . | | 41 |
| | 5 li | ntroduction | n à la méta-analyse | 42 |
| | 6 L | es données | pour la méta-analyse | 44 |

| | 7 | Plan d'a | analyses pour la méta-analyse des données transcriptomiques | 47 |
|--------|-------|------------|---|--------------|
| | 7. | 1 Métl | nodes et outils pour identifier les jeux de données et extraire les données des étu | ıdes 52 |
| | | 7.1.1 | Identifier les jeux de données | 52 |
| | | 7.1.2 | Les biais de publication | 53 |
| | | 7.1.3 | Extraction des données | 54 |
| | 7. | 2 Prép | arer, catégoriser et mettre à jour les jeux de données | 55 |
| | | 7.2.1 | Spécificité des données transcriptomiques en vue de leur méta-analyse | 55 |
| | | 7.2.2 | La normalisation | 56 |
| | | 7.2.3 | Mise à jour des identifiants | 57 |
| | 7. | 3 Choi | sir une méthode de méta-analyse pour identifier des différentiels d'abondance d | e gène |
| | | 61 | | |
| | | 7.3.1 | Fusion directe | 63 |
| | | 7.3.2 | Combinaison de rang | 63 |
| | | 7.3.3 | La taille de l'effet | 64 |
| | | 7.3.4 | Combinaison de taille de l'effet | 65 |
| | | 7.3.4 | 1 Effet fixe de la combinaison de taille de l'effet (FEM) | 65 |
| | | 7.3.4 | 2 Effets aléatoires de la combinaison de taille de l'effet (REM) | 66 |
| | | 7.3.5 | Les valeurs p | 67 |
| | | 7.3.6 | Combinaison de valeur p | 68 |
| | | 7.3.7 | Taille de l'effet versus valeurs p | 69 |
| | 7. | 4 Anal | yser, présenter et interpréter les résultats | 70 |
| | | 7.4.1 | Ontologie de gènes et enrichissement | 71 |
| | | 7.4.2 | Co-expression des gènes | 71 |
| | | 7.4.3 | Réseaux moléculaires | 72 |
| | | 7.4.4 | Sécrétion des protéines | 74 |
| | 8 | Outils d | e méta-analyse pour les données de transcriptomique | 76 |
| | Obj | ectifs | de thèse | 79 |
| | Rés | ultats | | 83 |
| | Poin | ts clés de | la synthèse bibliographique : article 1 | |
| | Artic | le 1 : Pro | teomics Research in the Adipose Tissue | 86 |
| | Poin | ts clés de | l'article 2: | 110 |
| | Artic | le 2 : Pat | hways and biomarkers of marbling and carcass fat deposition in bovine reveale | d by a |
| combin | ation | of gel-ba | sed and gel-free proteomic analyses | 110 <u>2</u> |
| | Point | ts clés de | l'article 3: | 122 |

| Article 3 : Molecular signatures of muscle growth and composition deciphered by the meta-analysis |
|---|
| of age-related public transcriptomics data |
| Discussion, conclusions et perspectives |
| 9 Discussion générale 154 |
| 9.1 Les données protéomiques pour la compréhension de la composition corporelle et |
| musculaire des bovins |
| 9.2 La réutilisation des données publiques via la méta-analyse |
| 9.3 Les données omiques combinées, un atout pour comprendre L'Adiposité Musculaire ? 162 |
| 10 Conclusions et perspectives163 |
| Annexes |
| Poster |
| Points clés du script d'analyse des résultats de proteomique : Management of proteomic data to |
| identify differentially abundant proteins according to one discriminant factor and to correlate their |
| abundance with the value of this factor |
| Tableau des données170 |
| Bibliographie |

Table des Figures

| Figure 1 : A: Note de la conformation (de E pour Excellent à P pour médiocre) selon la grille E.U.R.O.P. B: | |
|--|----|
| Note de l'engraissement, de 1 pour une carcasse avec peu de gras sous-cutané à 5 pour une carcasse | |
| présentant beaucoup de gras sous-cutané, selon a grille E.U.R.O.P 1 | 0 |
| Figure 2 : Courbe de croissance sigmoïde décrivant la croissance des bovins (Al-Jammas, 2017) | 4 |
| Figure 3 : Composition du croît lors de la croissance de bovins Frison. Adapté de (Robelin, 1986) | 5 |
| Figure 4 : Étapes majeures de la myogenèse chez le bovin. Provenant de Bonnet et al. (2010) 1 | 6 |
| Figure 5 : Croissance des différents TA chez le bovin (Bonnet et al., 2010). BAT : (pour Brown Adipose | |
| Tissue), TA bruns et WAT : (pour White adipose tissue) tissu adipeux blanc. Les TA beiges ne sont pas | |
| représentés, car peu connus, non-caractérisés chez le bovin 1 | 8 |
| Figure 6 : Différence de composition de carcasse pour quatre races bovines à âges similaires. Adapté de Gotoh | |
| et al. (2009) | 0 |
| Figure 7 : Différence de répartition des TA pour quatre races bovines à âges similaires. Adapté de Gotoh et al. | |
| (2009) | 1 |
| Figure 8 : Relation souvent linéaire entre le dépôt adipeux intramusculaire et le poids de la carcasse provenant | |
| de Gotoh et al. (2009). IMF : lipide intramusculaire ; weight : poids de carcasse | 3 |
| Figure 9: Schéma de fonctionnement des puces à ADN provenant de Duggan, Bittner, Chen, Meltzer et Trent | |
| (1999) | 0 |
| Figure 10 : Schéma du fonctionnement de la fixation des ARNm lors d'un séquençage ARNm (RNA-seq) 3 | 1 |
| Figure 11 : Schéma de récupération des billes et fonctionnement de la reverse transcriptase lors d'un séquençag | е |
| d'ARN | 2 |
| Figure 12: Diagramme de Venn des gènes différentiellement exprimés décrits dans le Tableau 7 | 9 |
| Figure 13: Nombre d'articles par année faisant référence à "meta-analysis" ou "systematic review" dans PubMe | d |
| (1948-2019). Recherche faite le 11/09/2019 | .3 |
| Figure 14 : Diagramme des types de données et de méthodes classiquement utilisées en méta-analyse 4 | 5 |
| Figure 15: Description du plan d'analyse de la méta-analyse de Farhadian et al 2018 4 | .9 |
| Figure 16: Description du plan d'analyse proposé par Rung et Brazma 2013 4 | .9 |
| Figure 17 : Lien entre les 4 bases de dépôt de données (source : https://bioinfo-fr.net/bases-de-donnees-de- | |
| sequencage-geo-sra-ena-arrayexpress) | 3 |
| Figure 18 : Funnel plots montrant la relation entre la taille de l'effet et la taille des échantillons. A : en l'absence | e |
| de biais de publication, le nuage de point attendu a une forme en entonnoir avec des points répartis | |
| symétriquement. B : un profil asymétrique du nuage de point suggère un biais de publications, et il n'y a | |
| que peu d'études négatives à faible effectif (figure provenant de (Dawson et al., 2016))5 | 4 |
| Figure 19 : Étapes d'obtention de la séquence nucléique d'un gène dans la base de données GEO sur la | |
| plateforme GPL4196 pour un utilisateur lambda5 | 9 |
| Figure 20 : Objectifs des méta-analyses sur des données de puces à ADN déclarés dans 328 publications (adapt | é |
| de Tseng et al. 2012). Quatre types d'objectifs, avec des proportions (en %) variables, ont été catégorisés, | |
| le majeur étant la recherche de gènes différentiellement exprimés ϵ | 2 |

| Figure 21 : Méthode de méta-analyses utilisées dans les 191 méta-analyses cherchant à mettre en évidence les |
|--|
| gènes différentiellement exprimés (adaptée de Tseng et al. 2012) |
| Figure 22 : les 25 gènes les plus surexprimés. Le rang est celui observé pour chacune des 5 études analysées par |
| DeConde et al., (2006) |
| Figure 23 : Illustration de la manière de calculer l'effet fixe de la taille de l'effet par Borenstein et al. (2010). La |
| taille de l'effet de chaque étude est représentée par un carré et par l'intervalle de confiance |
| Figure 24 : Illustration de la manière de calculer l'effet aléatoire de la taille de l'effet par Borenstein et al. |
| (2010). La taille de l'effet de chaque étude est représenté par un carré et par l'intervalle de confiance 67 |
| Figure 25: Comparaison de 3 publications portant sur les différences de protéome entre des dépôts adipeux chez |
| un même animal. DA correspond à Différentiellement Abondantes |
| Figure 26: Comparaison de 5 publications portant sur la différence de protéome entre des animaux différents par |
| le persillé. DA correspond à Différentiellement Abondantes |
| Figure 27 : Résumé graphique des résultats de l'article. Les protéines qui ont une abondance positivement |
| corrélée aux dépôts adipeux de la carcasse ou du muscle sont en vert tandis que celles qui présentent une |
| corrélation négative sont en rouge. Les protéines qui ont une relation avec le pourcentage de tissu adipeux |
| de la carcasse, du muscle ou les 2 sont indiquées |
| Figure 28 : Schéma de sélection et d'analyse des données publiques transcriptomiques issues du muscle bovin |
| |
| Figure 29: Diagramme de Venn des protéines dont l'abondance varie avec la teneur en LIM. Produit en utilisant |
| http://jvenn.toulouse.inra.fr/app/example.html (Bardou, Mariette, Escudié, Djemiel et Klopp, 2014). Les |
| protéines utilisées de (Bazile, Picard, Chambon, Valais et Bonnet, 2019) sont uniquement celles corrélées |
| au dépôt adipeux intramusculaire. Pour rappel, l'étude Ceciliani est la synthèse des 5 publications de |
| protéomiques disponible jusqu'en 2017 |
| Figure 30: Diagramme de Venn des gènes différentiellement exprimés et communs entre les études rapportant la |
| différence de transcriptome selon l'âge des bovins. Les résultats de Moisá et al. (2013) et de Qin et al. |
| (2011) n'ont pas pu être inclus dans la comparaison en raison respectivement de l'absence de liste de gènes |
| DEG, et d'ID obsolète. Le faible consensus que nous observons entre les études provient probablement |
| d'une différence d'ID associés à chaque publication, et du faible nombre d'ARNm listé dans les |
| publications relativement aux données déposées dans GEO 160 |
| Figure 31 : Diagramme de Venn des protéines différentiellement abondantes selon la teneur en lipides et des |
| gènes différentiellement exprimés selon l'âge identifiés par la méta-analyse. Produit en utilisant |
| http://jvenn.toulouse.inra.fr/app/example.html (Bardou et al., 2014) et modifié en ajoutant les noms des |
| gènes/protéines communes aux données transcriptomiques et protéomiques. Les protéines utilisées de |
| (Bazile et al., 2019)(nommées Bazile_protéines) sont uniquement celles corrélées au dépôt adipeux |
| intramusculaire |

Table des tableaux

| Tableau 1 : La consommation de viande bovine en France. (1) Production indigène brute (2) Consommation |
|---|
| indigène brute. Source Agreste (2019) |
| Tableau 2 : Catégories de gros bovins et codes associés (Institut de l'élevage, 2007 ; Règlement UE |
| n°1308/2013) |
| Tableau 3 : Les outils de prédiction des qualités de la viande bovine. Synthèse adaptée d'une présentation de |
| Linda Farmer, 2016 (http://www.eaap.org/Annual_Meeting/2016_belfast/S05_06_Farmer.pdf) 12 |
| Tableau 4 : Pourcentage de Lipides intramusculaires (LIM) pour différentes races. n nombre d'animaux pour |
| chaque race ; ()- Ecart type; a,b: Dans une ligne, les valeurs ne partageant pas un exposant commun sont |
| significativement différentes pour la race. Adapté de Cafferky et al. (2019) |
| Tableau 5 : Pourcentage de lipide intramusculaire (LIM) selon les races et selon le sexe des animaux. ET : Ecart |
| Type. LT : Longissimus Thoracis, LD : Longissimus Dorsi. LL : Longissimus Lumborum. KC : Korean |
| cattle. Adapté de (Park, Beak, et al., 2018) |
| Tableau 6 : Nombre de publications protéomiques et de publications ou données transcriptomiques disponibles |
| par tissu et facteur de variation de la croissance des tissus adipeux et des muscles chez le bovin. TA : tissu |
| adipeux, TM : tissu musculaire. Pour les données transcriptomiques, le premier chiffre correspond au |
| nombre d'étude et le deuxième correspond au nombre de jeux de données déposés dans les bases publiques |
| sous forme de GSE. Un GSE peut contenir plusieurs GSM et donc des abondances de gènes selon |
| différents traitements |
| Tableau 7 : Gènes différentiellement exprimés entre des animaux divergents par la teneur en lipides |
| intramusculaires (LIM) dans 5 études |
| Tableau 8 : Les différentes étapes d'une méta-analyse conduite sur des études de puces à ADN (adapté de |
| Ramasamy et al., (2008)) |
| Tableau 9 : Liste non exhaustive d'outils pour la mise à jour des identifiants de gènes à partir de séquences |
| (tableau adapté de Siezen & Van Hijum (2010)) |
| Tableau 10 : Principaux modèles de combinaison de valeurs p 69 |
| Tableau 11 : Principales bases de données (non spécifique à un organisme) des interactions entre molécules 73 |
| Tableau 12 : Exemple d'outils de prédiction des protéines sécrétées, adapté de (Caccia, Dugo, Callari et |
| Bongarzone, 2013) |
| Tableau 13 : Principaux paquets R existant pour la méta-analyse de données transcriptomiques et leurs |
| méthodes statistiques. Les paquets en orange sont ceux qui ne sont plus disponibles sur le CRAN77 |
| Tableau 14 : Listes des 50 protéines uniques dont l'abondance varie selon la teneur en lipides intramusculaires. |
| Cette liste a été produite à partir des 5 publications disponibles début 2017. Les protéines trouvées dans au |
| moins 2 publications sont surlignées en gris |
| Tableau 15: Relation entre la variation de l'abondance d'une protéine et la variation de teneur en lipides |
| intramusculaires selon l'âge (Kim et al. 2009 ¹ ; Zhang et al. 2010 ²), la race (Keady et al. 2013 ³), ou le score |
| de persillage à 24 mois (Shen et al. 2012 ⁴) et 30 mois (Mao et al. 2016 ⁵) chez les bovins |
| Tableau 16: Parmi les protéines comparées dans la figure X, 9 ont été identifiées dans au moins 3 publications. |
| Le sens de variation de l'abondance des protéines est reporté en fonction de l'augmentation de la teneur en |

| lipides intramusculaires. Les 3 protéines surlignées en vert sont parmi les 9 identifiées dans au moins 2 de | es |
|--|----|
| 5 études utilisées dans Ceciliani et al. (2017) | 57 |
| Tableau 17: exemple de tableau avec les valeurs de corrélations pour un jeu test | 69 |
| Tableau 18 : Description des paramètres animaux liés au log ratio (comparaison de 2 conditions) des jeux de | |
| données transcriptomiques | 70 |

LISTE DES ABBREVIATIONS

A

AD : Données résumés (Aggregate Data)

ADN : Acide DesoxyriboNucléique

ARN : Acide RiboNucléique

ARNm : ARN messager

ADNc : ADN complémentaire

C

CRAN: Comprehensive R Archive Network

D

DE: Différentiellement exprimé

DEG: Gène(s) différentiellement exprimé(s) differentially expressed gene(s))

DIP: Données individuelle de participant (voir IPD)

DA : Données Agrégées / Données Résumées (voir AD)

dNTP : désoxy-Nucléotides tri-phosphate

G

GEO : Gene Expression Omnibus

GO : Ontologie des gènes (Gene Ontology)

GPL : GEO Plateform (GEO PLateform)

GSE : GEO Series Experiment

GSM : GEO Sample (GEO SaMple)

GAQ : GO Annotation Quality (see QAG)

Ι

IPD: Données indivduelles de participants (Individual participant Data)

L

Lowess (methode de normalisation): Locally weighted scatter plot smoothing

LIM : lipides intra-musculaire

Μ

MSA : Meat Standard Australia

MS : Spectrométrie de masse

P

P-valeur : valeur de probabilité (probability value)

PCR: Polymerase chain Reaction, également nommé ACP, amplification en chaîne par

polymérase

Pb : paire de base

PPi : interaction protéine-protéine

Q

QAG : Qualité d'Annotation de l'Ontologie de Gene

qPCR : PCR quantitative

R

RNA-seq : RNA-sequencing, séquençage ARN

Avant-propos

AVANT-PROPOS

Avant-propos

Ce dernier siècle se caractérise par deux grandes révolutions qui ont impacté la pratique de la recherche : la révolution du Big Data qui est la production très importante par les scientifiques de données et l'Open Data, qui consiste en la disponibilité et l'accessibilité de ces données via notamment internet. Avec le développement des méthodes de production de données telles que les technologies de puces à ADN ou de RNAseq qui se démocratisent, le volume de données relatives à l'expression du génome augmente. Ainsi des bases de données telles que celle de l'Institut Européen de Bioinformatique (EBI), une des bases de données biologiques les plus importantes avait approximativement 155 petabytes de données sur les gènes, protéines ou petites molécules en 2017 contre 101 petabytes en 2016 (« EBI Scientific Report 2017 Contents », 2017). Ces bases de données publiques telles que celle de l'EBI, mais également de Web Of Science, GEOdataset ou proteomeXchange, permettent aux scientifiques de déposer leurs données et de les partager. Ces bases de données contiennent des données relatives à l'expression du génome, produites par des méthodes de quantification de l'abondance de l'ensemble des ARNm (méthodes de transcriptomiques) ou de protéines (méthodes de protéomiques) présents à un instant t dans un type cellulaire ou un tissu.

Pour les filières de production bovine, deux critères déterminent la valeur économique des carcasses et les qualités nutritionnelle et sensorielle des viandes. La composition corporelle, correspond à la proportion de muscle, tissu adipeux (TA) et os dans une carcasse et détermine le rendement en muscle/viande extrait d'une carcasse. La composition tissulaire du muscle aussi nommée composition musculaire dans la suite du manuscrit correspond à la proportion de lipides relativement aux protéines dans le muscle. La valeur économique de la carcasse d'un animal résulte donc de l'équilibre entre les masses de tissus musculaire et adipeux qui la composent, et les qualités d'un morceau de viande résultent en partie de l'équilibre entre cellule musculaires contractiles et adipocytes présentes dans un muscle. Ces 20 dernières années de nombreuses données transcriptomiques et protéomiques appliquées aux muscles et TA de bovins ont décrit les mécanismes moléculaires qui régulent la croissance de chacun des tissus et l'équilibre muscle/TA en fonction de l'âge, de la race, du sexe ou des pratiques liés à l'élevage des bovins. Cependant, la synthèse et la généricité des mécanismes identifiés restent difficiles à établir, du fait du volume de données, et de la diversité des facteurs de variations ou des méthodologies appliquées. Des études récentes soulignent que seulement quelques pourcents des résultats d'études protéomiques (Vaudel et al., 2016) et transcriptomiques (Rung et Brazma, 2013) sont exploités par les auteurs et que la réutilisation et la ré-analyse de ces données produisent de nouvelles connaissances. Une méthode de combinaison des données qu'elles soient transcriptomiques ou protéomiques est la méta-analyse, une démarche statistique combinant les résultats d'une série d'études indépendantes afin d'en tirer une conclusion générale.

Dans ce contexte de big data et d'open data, l'objectif général de ma thèse est de combiner et d'analyser les données protéomiques et transcriptomiques disponibles chez les bovins producteurs de viande afin de mettre en évidence des mécanismes et des biomarqueurs de la composition corporelle ou tissulaire.

Aperçu de la thèse

La thèse est divisée en 4 grandes parties. La 1ère partie est une synthèse bibliographique contextualisant la production et la qualité des produits bovins orientés viandes, les outils d'évaluations de la carcasse et de la viande en abattoir et des facteurs influençant la composition corporelle ou musculaire des bovins. Un 2ième chapitre liste les données transcriptomiques et protéomiques disponibles et produites afin de comprendre la croissance des TA ou musculaire des bovins selon différents facteur d'élevage. Le 3ème chapitre décrit les méthodes de méta-analyses appliquées aux données omiques, les modèles utilisés ainsi que les différentes étapes et outils pour mener une méta-analyse.

La 2^e partie de la thèse présente les objectifs de la thèse.

La 3^e partie présente les résultats de mes travaux de recherche. Le premier article est un chapitre d'ouvrage qui synthétise les connaissances disponibles en lien avec la croissance des TA et la composition musculaire des bovins. Le deuxième article apporte des données protéomiques qui complètent le manque de connaissance des indicateurs de la composition corporelle des bovins de race plus tardive. Le troisième article rapporte les mécanismes identifiés et proposés comme contribuant à l'équilibre cellules musculaires et cellules adipeuses dans le muscle, grâce à la méta-analyse de données transcriptomiques ayant comme facteur de variation majeur l'âge.

La 4ième et dernière partie de cette thèse est composée d'une discussion et d'une conclusion générale, avec les perspectives qu'offrent la réalisation de ce travail.

CHAPITRE 1 : PRODUCTIONS ET QUALITES DES CARCASSES ET VIANDES BOVINE

1 PRODUCTION ET QUALITE DES PRODUITS DES BOVINS

1.1 **PRODUCTION ET CONSOMMATION DE LA VIANDE BOVINE EN FRANCE**

La France est le premier producteur et consommateur de viande bovine en Europe: 1.55 million de tonnes/an (« La consommation de viande en France », 2018) et 23.1 kg consommés/habitant/an en 2018. La France a exporté environ 200 000 tonnes de viande bovine dans le monde en 2017.

La viande bovine consommée en France est issue en majorité de femelles provenant du cheptel allaitant (c'est-à-dire de races à viande ou mixte (lait et viande) destinées à l'élevage de veaux pour la production de viande), le cheptel allaitant représentant 2/3 de la viande française abattue (Laurioux, 2018). L'Auvergne-Rhône-Alpes fait partie des 5 régions qui cumulent à elles seules 75% de l'effectif total de vaches allaitantes (essentiellement issues de race à viandes) aux côtés de la Nouvelle-Aquitaine, l'Occitanie, la Bourgogne-Franche-Comté et le Pays de la Loire (FRANCEAGRIMER, 2018).

L'offre de viande bovine en France est très hétérogène, car elle est issue de bovins mâles (entier ou castré) ou femelles, d'âge et de race variables, élevés pour la production de lait (race laitière), viande (race à viande) ou les deux (Laurioux, 2018). En effet au sein d'un même type racial, la viande bovine provient de différents types d'animaux âgés au minimum de 8 mois : les génisses (femelles âgées de plus de 12 mois et nullipares), les vaches (au moins primipares), les jeunes bovins ou taurillons (mâles non castrés entre 12 et 24 mois d'âge), les bœufs (mâles castrés et de plus de 12 mois) et les taureaux (mâles non castrés de plus de 24 mois). La diversité de la production bovine française provient aussi de la multiplicité des races, la France étant un pays privilégiant l'élevage de race pure (pas moins de 25 races dénombrées sur le territoire français, sans compter les races importées ou les croisements (« Les races bovines »)). Enfin la diversité résulte aussi de la diversité des modèles d'élevage pour chaque type de bovin, lait ou viande.

En 2018, malgré un contexte de baisse de la consommation de viande bovine, la production bovine totale a augmenté de 0,9% (Agreste, 2019).

Tableau 1: La consommation de viande bovine en France. (1) Production indigène brute (2) Consommation indigène brute. Source Agreste (2019).

| Tonnes-équivalent-carcasse | Moyenne 2013-2017 | 2017 | 2018 |
|--|-------------------|---------|---------|
| Abattages totaux (a) | 1 436,5 | 1 442,2 | 1 460,0 |
| Exportations tous animaux vivants (b) | 221,5 | 235,6 | 234,3 |
| Importations tous animaux vivants ('c) | 6,1 | 4,4 | 6,3 |
| Production bovine totale (1) (a+b-c) | 1 652,0 | 1 673,3 | 1 687,9 |
| Exportations Viande (d) | 257,6 | 236,0 | 238,0 |
| Importations Viande (e) | 366,4 | 324,6 | 324,0 |
| Consommation totale (2) (a+e-d) | 1 573,4 | 1 552,8 | 1 528,1 |

La baisse régulière de consommation de viande provient de plusieurs facteurs (Laurioux, 2018) :

- La diminution de la consommation hebdomadaire du fait des préoccupations nutritionnelles, écologiques ou relatives au bien-être animal qui sont fortement relayées par les médias,
- Une diminution des connaissances des morceaux de viande et de leur mode de cuisson à l'origine de déception gustative,
- Un prix souvent jugé élevé pour les consommateurs qui ne s'y retrouvent pas toujours en termes de qualité gustative.

Enfin, la commercialisation de la viande bovine a largement évolué en France ces dernières années. Elle a lieu essentiellement via les grandes et moyennes surfaces, au détriment des boucheries traditionnelles/de détails qui ne représentaient plus que 13% des volumes de viande de boucherie fraiche achetée par les ménages en 2015 (KantarWorldPanel). Ce mode de commercialisation a vu émerger de nouveaux modes de préparation de la viande, notamment avec la mise sous vide en format individuel de viande pour une vente en libre-service. Il a aussi modifié les exigences pour l'emballage notamment ainsi que le cahier des charges des abattoirs et ateliers de découpe qui est de plus en plus précis quant à la conformation ou au poids de la carcasse. La logique industrielle s'est répandue dans les abattoirs ou les activités de désossage, de découpe et la fabrication de produit unitaire se sont imposées, la rentabilité de ces activités ne dépendant que du prix des carcasses bovines (Laurioux, 2018). Ceci explique les enjeux actuels pour optimiser par les pratiques d'élevage la production de carcasses et donc leur valeur économique.

1.2 OUTILS D'EVALUATION DE LA QUALITE DES CARCASSES A L'ABATTOIR

La qualité d'une carcasse de bovin dépend de son poids, de sa conformation c'est-à-dire de la taille et de la répartition des muscles dans les différentes régions anatomiques et enfin de son état d'engraissement et de la localisation des dépôts adipeux. À poids de carcasse identique, les indicateurs d'une carcasse de qualité supérieure sont une teneur optimale en gras et un rapport muscles/os élevé. La qualité gustative de la viande qui en résulte est, elle, tributaire à la fois du développement musculaire et adipeux dans le muscle. Le tissu adipeux (TA) apporte la jutosité et la flaveur de la viande. En France, la viande consommée (hors steak haché) contient en moyenne 3 à 8 % de lipides contre environ 15% pour les pays asiatiques.

En France et dans le monde, la qualité des carcasses est définie par une grille prenant en compte des indicateurs différents selon les pays. Ces grilles de classement s'appellent MSA pour Meat Standards Australia pour l'Australie, USDA scale (U.S Department of Agriculture) pour les États-Unis, EUROP pour l'Europe, Canada pour le Canada, Korea pour la corée du sud ou encore JMGA (Japanese Meat Grading Association) pour le Japon, et elles fixent le prix des carcasses payées aux producteurs. Les critères utilisés par chaque grille sont présentés dans Polkinghorne et Thompson (2010). Ils sont essentiellement visuels, diffèrent peu entre eux sauf en Europe et en Afrique du Sud ou le persillé n'est pas pris en compte ni aucun indicateur après refroidissement de la carcasse (Polkinghorne et Thompson, 2010).

Je présenterai la grille de notation E.U.R.O.P. utilisée en France, les 5 autres grilles étant présentées dans les synthèses de Polkinghorne and Thompson (2010) et Laurioux (2018).

En France, pour les gros bovins (poids vif >300kg, correspondant à un animal de plus de 8 mois environ), une première grille de classement a été mise en place dans les années 60 : c'était le catalogue F.R.A.N.C.E. Une grille de classement a été ensuite créée, la grille E.U.R.O.P.A, rendue obligatoire en France en 1977 pour les bovins. La grille utilisée aujourd'hui, la grille E.U.R.O.P, mise en place en 1981 pour tous les états membres de l'Union européenne s'est largement inspirée de cette grille. La grille EUROP avait pour vocation d'harmoniser les cotations dans tous les pays de l'Union européenne (AgriMer, 2010). La démarche pour classifier et évaluer la valeur économique d'une carcasse dans un abattoir français commence par la pesée de la carcasse selon les règles imposées par le décret n° 94-808 du 12 septembre 1994 et l'arrêté du 26 novembre 2000. La pesée est faite lors du classement EUROP, en fin de chaîne d'abattage sur la carcasse chaude. 2% de ce poids sont retirés forfaitairement du poids mesuré pour supprimer le poids lié à la quantité d'eau qui s'évapore de la carcasse au cours de son refroidissement. Ce nouveau poids correspond au poids de carcasse froide qui fixe la valeur de la carcasse. La grille EUROP considère les typologies animales, l'état d'engraissement et la conformation de la carcasse. Les typologies animales sont la catégorie communautaire et la dénomination commerciale, désignée par une lettre selon l'âge et le sexe du bovin (cf Tableau 2). La grille EUROP, note d'abord la conformation de la carcasse et donc la masse et la répartition anatomique des muscles selon 5 lettres de E pour excellente à P pour médiocre (cf Figure 1A). Puis elle note l'engraissement visuel de surface de 1 pour très peu gras (engraissement qui est visible avec le gras sous-cutané) à 5 pour une carcasse très grasse (cf Figure 1B). Le prix moyen de race à viande de conformation E était de 4.40€/kg net et de 3.75€/kg net pour une conformation U au marché de Bourg-en-Bresse, pour des génisses toutes races à viande confondues en date du 24/09/2019 (TERRE-NET, s. d.). Ce prix au kilo multiplié par le poids de la carcasse détermine la valeur économique de la carcasse.

Tableau 2 : Catégories de gros bovins et codes associés (Institut de l'élevage, 2007 ; Règlement UE n°1308/2013)

| Catégories communautaires | А | Mâles non castrés de moins de 2 ans |
|------------------------------|------------|---|
| | В | Autres mâles non castrés (taureaux) |
| | С | Mâles castrés (bœufs) |
| | D | Femelles ayant déjà vêlé (vaches) |
| | E | Autres femelles (génisses) |
| Dénomination commerciale | Z A Z E | mâles (Z A) ou femelles (Z E) de plus de 8 mois et de 12 mois maximum |

Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines



Figure 1 : A: Note de la conformation (de E pour Excellent à P pour médiocre) selon la grille E.U.R.O.P. B: Note de l'engraissement, de 1 pour une carcasse avec peu de gras sous-cutané à 5 pour une carcasse présentant beaucoup de gras sous-cutané, selon a grille E.U.R.O.P

Le système de classement communautaire conditionne donc la rémunération à l'éleveur, cependant pour un même classement de conformation et d'état d'engraissement cachent une grande variabilité individuelle de rendement en viande nette commercialisable.

D'autre part, des travaux ont montré un écart entre les critères de la grille EUROP et les méthodes de référence utilisées en expérimentation pour estimer la qualité de la carcasse, c'està-dire la proportion d'os, de muscle et de gras. En effet en conditions expérimentales, les méthodes de référence pour la qualité des carcasses sont la dissection totale pour mesurer directement les compositions tissulaires ou chimiques de la carcasse. Une méthode indirecte est la dissection des os, muscle et gras de l'une des côtes à partir de laquelle les poids sont utilisés pour calculer la proportion de chaque tissu dans la carcasse grâce à des équations établies à partir des dissections de carcasses (Gagaoua, Picard et Monteils, 2018; Robelin, Geay, Jailler et Cuylle, 1975). Drennan, McGee et Keane (2008) ont montré que les notes de conformation et état d'engraissement expliqueraient 74% et 57% de la variation de la proportion de TA respectivement chez les mâles et les génisses. Le poids de la carcasse seul, le poids de la carcasse avec la note de conformation et le poids de la carcasse avec les notes de conformation et l'état d'engraissement expliqueraient respectivement 0,1%, 37,9% et 46,7% de la variation totale du rendement de la viande commercialisable (Perry, Yeates et McKiernan, 1993). Aussi, en 2017, Monteils et al. (2017) ont proposé d'ajouter différents indicateurs à la classification EUROP, afin d'améliorer l'évaluation d'une carcasse bovine. Ces nouveaux critères incluaient notamment le score de persillé, ou gras intramusculaire, du faux-filet, l'un des critères utilisés par la grille américaine USDA. Enfin, que ce soit pour le système EUROP ou le système MSA (Bonny et al., 2016) qui inclue pourtant plus de critères dans la cotation, aucun ne prédit la qualité de la viande et des outils de classification spécifiques à l'évaluation de la qualité des viandes tentent d'être développés par et pour les industriels de la viande.

La classification EUROP est faite soit par des classificateurs, soit par des machines à classer, ayant l'agrément délivré par FranceAgriMer, seul organisme autorisé à évaluer leur performance également. Le classement européen se fait dans l'abattoir dans l'heure suivant l'abattage de l'animal. La catégorie et le classement sont estampillés sur les carcasses, de manière indélébile. Dans le cas de l'utilisation d'une machine, le choix définitif du classement revient toujours à l'opérateur. La classification des carcasses se fait donc uniquement sur des appréciations visuelles, quels que soient les systèmes de classification utilisés, que ce soit fait par des machines ou par des experts. Pour pallier le manque d'outil précis de composition de la carcasse, différentes méthodes commencent à être proposées. Parmi celles-ci, des équations de prédiction de la composition en muscle et en tissus adipeux à partir du poids de carcasse chaud, du type d'animal (précoce ou tardif pour le dépôt adipeux) ou encore l'épaisseur de gras sous-cutané ont été mises en place (Al-Jammas, 2017).

1.3 Les outils d'evaluation de la qualite de la viande

Une forte demande émane des industriels pour un développement d'outils de prédiction des qualités de la viande : qualités nutritionnelles (macro et micronutriments, teneur en lipides, et composition en acides gras) et organoleptiques (tendreté, jutosité et flaveur). Ces qualités sont largement dépendantes de la composition cellulaire des muscles, c'est-à-dire de la proportion de fibres musculaires, cellules conjonctives, adipeuses, et dans une moindre mesure, vasculaires et nerveuses. Dans ce paragraphe je présente les méthodes qui estiment la teneur en lipides ou la proportion muscle/gras d'une pièce de viande.
Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines

Différents outils de prédiction de paramètres liés aux qualités de la viande ont déjà été testés avec plus ou moins de succès. Les différentes méthodes décrites sont les suivantes (Farmer et Farrell, 2018) :

| Tableau 3 : Les | outils de prédiction d | les qualités de la | viande bovine. Synthèse | adaptée d'une |
|----------------------------|------------------------|--------------------|-------------------------|---------------|
| présentation | de | Linda | Farmer, | 2016 |
| (<u>http://www.eaap</u> . | org/Annual_Meeting/20 |)16_belfast/S05_06 | <i>_Farmer.pdf</i>). | |

| Instrument de mesures | Critères de qualités de viande prédits | Références |
|---|---|--|
| Imagerie hyperspectrale (HSI) | •Tendreté de la viande •Présence de microbes | (ElMasry, Kamruzzaman, Sun et Allen, 2012) |
| | viande | (Peng et Dhakal, 2015) |
| Spectroscopie infrarouge et proche infrarouge (SPIR) (Vis-NIR) | Persillé dans le <i>longissimus lumborum</i> Composition en acides gras. | (Yancey, Apple, Meullenet et Sawyer, 2010) (Peng et Dhakal, 2015) (Mourot et al., 2014) |
| Spectroscopie Raman | Force de cisaillement Warner-Bratzler.Persillé et composition en acides gras | (Roehe et al., 2013) |
| Robot à pH | •Qualité de la carcasse pendant le refroidissement (Utilisé dans la grille MSA). | (Polkinghorne et Thompson, 2010; Roehe et al., 2013). |
| Scanner de tomographie informatique (CT) | •Composition en gras, muscle et os de l'animal de son vivant ou de la carcasse. Outil encombrant et très coûteux. | Font-i-Furnols et al., 2014 (Font-I-Furnols, Carabús, Pomar et Gispert, 2014) |
| Imagerie à Résonnance magnétique (MRI) | •Persillé d'un morceau de viande (outil coûteux et encombrant). | (Sangdae Lee et al., 2015) |
| Analyse d'image | Persillé d'un morceau de viande | (Peng et Dhakal, 2015) (Giaretta et al., 2018) |
| Ultrasons | Persillé d'un morceau de viandeEpaisseur de gras sur la carcasse | (Tait, 2016) |

Des différentes méthodes citées, beaucoup sont coûteuses (CT, IRM), difficilement utilisables en abattoir, en cours de développement pour une utilisation industrielle (analyse d'image) ou encore spécifiques à des types d'animaux (comme la technique de vision par ordinateur qui est spécifique aux animaux de 2 à 21 jours). La SPIR (Spectrométrie Proche Infra-Rouge) est une méthode rendue ambulante et donc utilisable en abattoir. Cette méthode a été utilisée pour la première fois pour prédire la composition en protéines et en gras de bœuf haché en 1996 (Isaksson, Nilsen, Tøgersen, Hammond et Hildrum, 1996). Cependant, l'utilisation sur un morceau tel que la 12^{ième} côte pose plusieurs problèmes. En effet, il a été montré qu'il y a un manque d'homogénéité de résultats pour un même échantillon, et que le gras de carcasse peut être un facteur de variation des résultats, notamment pour l'estimation du

persillage. Pour augmenter la qualité de prédiction, il faudrait augmenter la surface observée ce qui augmenterait le temps de scan pour chaque échantillon (Nuria Prieto, Pawluczyk, Dugan et Aalhus, 2017). De plus ces mêmes auteurs ont montré plus tôt que cette méthode n'était pas générique, car des différences de prédiction de profils d'acide gras étaient observées entre 2 races (Limousin et Aberdeen Angus) du fait de la variation de taille d'adipocytes, eux-mêmes influençant l'arrangement du dépôt adipeux intramusculaire (le persillage) au sein du muscle (N. Prieto et al., 2011). D'autres auteurs, en appliquant la méthodes SPIR à des échantillons de muscle broyé ont prédit avec un taux de succès de 95% pour le persillé (Mourot et al., 2014). Dans ce cas la méthode est destructive et utilise 50 à 100 g de muscle pour une bonne homogénéité de l'échantillon. La SPIR, bien que présentant une facilité d'utilisation, présente néanmoins des inconvénients, dont un manque de fiabilité des résultats d'une race à l'autre ou pour un même échantillon.

L'analyse d'image pour mesurer la quantité de lipides intramusculaire a été initiée en 2014 (Giaretta et al., 2018; Meunier et al., 2014). L'idée est de prendre en photo un morceau de viande et d'estimer la quantité de lipides intramusculaire visible, grâce à des algorithmes de traitement de l'image ou des réseaux de neurones. Une application pour smartphone, meat@ppli (<u>http://meatappli.idele.fr</u>), est en cours de développement et permet d'estimer la teneur en gras de la carcasse ou de la pièce de viande en temps réel à partir d'une photo.

1.4 LES FACTEURS DE VARIATIONS INFLUENÇANT LES ADIPOSITES CORPORELLE ET MUSCULAIRE DES BOVINS

En Europe, un minimum de 3-4% de lipides est nécessaire pour assurer le goût et une jutosité à la viande. Les lipides dans le muscle influencent aussi la tendreté et contribueraient à 10 à 15% de la variabilité de l'appréciation globale de la viande bovine par les consommateurs (Hocquette et al., 2010). La quantité et la qualité des lipides stockés dans le muscle résultent à la fois du nombre d'adipocytes présents dans les tissus et de l'équilibre entre synthèse et dégradation des triglycérides (Gondret et Hocquette, 2006).

La teneur en lipides de la viande bovine est très variable en fonction du morceau. Par exemple, en France, le tende de tranche (*Semimembranous*) et l'entrecôte (*Longissimus Thoracis*) issus de vaches de réforme contiennent respectivement 2 et 10 g de lipides/100 g de

viande (Bauchart, Chantelot et Gandemer, 2008). Il est classiquement admis que les muscles glycolytiques sont plus pauvres en lipides que les muscles oxydatifs et ceci, quelle que soit la race (Bonnet et al., 2004). Les facteurs de variation majeurs de l'adiposité corporelle ou musculaire des bovins, sont l'âge, la race, le sexe et la nutrition.

1.4.1 L'âge

Le facteur de variation le plus important de l'adiposité de la carcasse et du muscle est bien l'âge de l'animal. Les animaux, et notamment les bovins, suivent une courbe de croissance (accroissement de la taille et du poids) de type sigmoïde afin d'atteindre un poids adulte (Figure 2). Cette croissance connait 2 phases : une croissance accélérée, de la naissance au point d'inflexion ; et une phase de croissance ralentie, du point d'inflexion jusqu'à l'âge adulte.





Cependant, lors de la croissance, les tissus ne se développent pas tous à la même vitesse, comme en atteste le graphique de croissance postnatale des principaux tissus corporels (Figure 3). Schématiquement la croissance des TA augmente quand celle des muscles diminue.



Figure 3 : Composition du croît lors de la croissance de bovins Frison. Adapté de (Robelin, 1986).

1.4.1.1 Croissance du tissu musculaire

La croissance musculaire débute très tôt pendant la vie fœtale. La première population de cellule myogénique apparait 30 jours post-conception et elle se différentie en fibre de type I (lente, oxydative) à la fin du second trimestre de gestation soit 180 jours post-conception, la gestation chez le bovin étant d'environ 280 jours. Le second type de cellule myogénique prolifère entre 90 et 180 jours post-conception et se différencie en fibre de type II (rapide, glycolytique). Ainsi, le nombre de myofibres est fixé à la fin du second trimestre de gestation et les propriétés du muscle sont presque complètement établies avant la naissance du veau (Bonnet, Cassar-Malek, Chilliard et Picard, 2010; Chaze, Meunier, Chambon, Jurie et Picard, 2008) (Figure 4).

Le nombre de myofibres étant fixé à la naissance, la croissance postnatale du tissu musculaire résulte essentiellement de l'hypertrophie des fibres du fait de la synthèse et du dépôt de protéines et de la fusion des cellules satellites avec les fibres musculaire du fait de l'apport de cytoplasme et de noyau (Pallafacchina, Blaauw et Schiaffino, 2013).



Figure 4 : Étapes majeures de la myogenèse chez le bovin. Provenant de Bonnet et al. (2010).

1.4.1.2 Croissance du tissu adipeux

Le TA est un tissu diffus qui peut représenter de 5% à 35% de la masse corporelle (contre 55% de la masse corporelle correspondant à du muscle squelettique) du bovin selon l'âge, mais également le génotype (race et sexe) et la nutrition. Il y a 3 sites anatomiques distincts majeurs du dépôt adipeux : viscéral, sous-cutané et musculaire.

Chez les bovins, à la naissance les TA, excepté le TA sous-cutané, sont considérés comme des TA hétérogènes composés d'adipocytes blancs et d'adipocytes bruns (Bonnet et al., 2010; Taga et al., 2012). En effet chez la plupart des mammifères, il y a 3 types de TA : blanc, brun et beige, ce dernier étant un mélange de TA blanc et brun (Du, Wang, Fu, Yang et Zhu, 2015). Le TA brun est spécialisé dans le stockage et la mobilisation de lipides en vue de la thermogénèse, grâce à un grand nombre de mitochondries. Ce tissu est notamment caractérisé par l'expression de la protéine-1 de découplage (UCP1), présent dans la membrane interne de la mitochondrie. C'est le découplage entre la chaine respiratoire et la production d'ATP qui permet la production de chaleur, nécessaire à la survie des nouveau-nés ou des mammifères exposés au froid. Le TA blanc est spécialisé dans la synthèse et le stockage d'énergie sous forme de lipides. La croissance des TA, que ce soit brun, beige ou blanc résulte de la croissance du nombre d'adipocytes (hyperplasie) et de leur volume (hypertrophie).

Il est admis que le nombre d'adipocytes augmente essentiellement durant les phases précoces de la croissance des TA des bovins (Robelin, 1981; Robelin et Casteilla, 1990;

Vernon, Beard, Graves et Wilson, 1986). Il évoluerait peu par la suite et serait fixé dès l'enfance et l'adolescence chez l'Homme (Spalding et al., 2008) ou à 100 kg de poids vif chez le bovin (Robelin et Casteilla, 1990). Même si l'hyperplasie est considérée comme très faible chez les mammifères adultes, elle participerait cependant au renouvellement des adipocytes (Spalding et al., 2008) et à la croissance des TA quand l'organisme est exposé à une prise calorique ou une adiposité excessive. Ainsi, chez le bovin le nombre d'adipocytes augmente lorsque la taille moyenne des adipocytes atteint 90 µm (Garbutt, Anthony, Walker et McGuire, 1979; Robelin, 1981; Schoonmaker, Cecava, Fluharty, Zerby et Loerch, 2004) et ceci en réponse à des facteurs physiologiques tels que l'âge ou la nutrition. Par ailleurs, la croissance des TA tardifs musculaires résulte d'une augmentation du nombre d'adipocytes y compris durant la croissance post-natale.

L'hypertrophie est cependant le mécanisme principal de croissance des dépôts adipeux après la naissance du veau avec une croissance de 100 fois du volume des adipocytes dans les TA blanc périrénaux et de 4 à 10 fois dans les TA sous-cutanés, intermusculaires et omentales entre 150 et 600 jours de vie (Bonnet et al., 2010; Vernon et al., 1986). Cette hypertrophie résulte de la synthèse de lipides qui implique 3 voies lipogéniques : la lipogenèse *de novo* (synthétise des acides gras à partir d'acétate et de glucose), le captage des acides gras circulants sous forme de triglycérides associés au lipoprotéines ou sous forme libres, et enfin l'estérification des acides gras sous forme de triglycérides. Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines



Figure 5 : Croissance des différents TA chez le bovin (Bonnet et al., 2010). BAT : (pour Brown Adipose Tissue), TA bruns et WAT : (pour White adipose tissue) tissu adipeux blanc. Les TA beiges ne sont pas représentés, car peu connus, non-caractérisés chez le bovin.

1.4.1.3 Les compétitions lors de la croissance des tissus adipeux et musculaire

En fonction de l'âge, il y a une priorisation pour la croissance du muscle relativement à celle des TA (Figure 3). Un constat est que la croissance des TA a lieu au détriment de la croissance du muscle, comme cela est observé chez les races bovines extrêmes pour leur rapport masse grasse/masse maigre. Ainsi une compétition ou une priorisation pour la croissance des TA et musculaire surviendrait lors des phases de croissance par hyperplasie et lors des phases de croissance par hypertrophie.

Durant des phases précoces de croissance par hyperplasie, des relations ou des interactions entre la croissance musculaire et la croissance adipeuse sont fortement suggérées par des variations inverses et simultanées du nombre d'adipocytes et du nombre total de fibres

musculaires en réponse à des modifications de l'environnement intra-utérin, et notamment selon la nutrition maternelle et fœtale, ou selon le génotype des fœtus (Bonnet et al., 2010). Les connaissances actuelles suggèrent qu'au moins un précurseur commun ou deux précurseurs distincts des lignages adipeux et musculaires existent dans des proportions qui dépendraient de la localisation anatomique des TA, de l'espèce (en liaison avec la présence ou pas de TA brun), et/ou de l'environnement fœtal. Les régulations de la détermination de ces divers précurseurs participeraient alors à la balance masse de muscle/masse de TA au niveau du corps entier. En accord avec cette hypothèse, un précurseur commun aux fibres musculaires et aux adipocytes a été identifié chez la souris (Seale et al., 2008) et l'Homme (Pisani et al., 2011). Des précurseurs adipeux et musculaires distincts ont aussi été décrits chez la souris selon la localisation anatomique des TA (Seale et al., 2008; Y. H. Tseng et al., 2008).

Durant les phases tardives de la croissance, en particulier lors de la croissance par hypertrophie, une interaction négative entre tissu adipeux et musculaire pour l'utilisation des nutriments énergétiques expliquerait la croissance tardive des TA par rapport aux muscles. L'une des origines de cette interaction négative est une compétition pour l'utilisation des mêmes nutriments tels que les lipides circulants hydrolysés par la lipoproteines lipase fortement actives dans les 2 tissus (Bonnet et al., 2004). Au niveau du corps entier, un exemple frappant d'interaction négative est une croissance plus précoce des TA, au détriment de la croissance musculaire chez les génotypes peu musclés, et cette dynamique de croissance affecte de la même façon tous les TA y compris le TA intramusculaire (Gotoh et al., 2009; Picard, Lefaucheur, Berri et Duclos, 2002). En outre, chez des bovins ayant subi une restriction alimentaire, une croissance importante du muscle a été observée ainsi qu'une inhibition ou un retard de croissance du dépôt adipeux (Bonnet et al., 2010). Les mécanismes mis en jeux seraient des signaux endocrines/paracrines et/ou des interactions cellulaires physiques entre les types cellulaires (Bonnet et al., 2010). Ainsi, la protéine myostatine, impliquée dans la croissance du tissu musculaire, a été montrée comme inhibant l'expression de PPARy et C/EBPa, deux facteurs de transcription impliqués dans la différenciation des préadipocytes bovins (Hirai et al., 2007). Chez les ruminants sous-nutris, l'orientation des nutriments et des métabolismes est modifiée grâce à l'intégration de différents signaux tels que l'insuline, l'hormone de croissance (GH, growth hormone), le cortisol, les catécholamines, le glucose ou encore des adipokines telle que la leptine, qui présente une importante activité endocrine (Chilliard, Delavaud et Bonnet, 2005).

Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines

Or l'équilibre entre la croissance du TA et celle du tissu musculaire va être influencée par différents facteurs, l'âge, la race (Shingu et al., 2001), le sexe (Gettys, Henricks et Schanbacher, 1988) ou l'alimentation (Bonnet et al., 2010) et plus particulièrement le ratio GH/insuline.

1.4.2 La race

Le deuxième facteur de variation majeur influant les adiposités corporelle et musculaire est la race de bovin. Il existe des différences selon la race, que ce soit pour la quantité relative de TA dans la carcasse à âge égal (Figure 6) ou pour la répartition anatomique des dépôts adipeux (Figure 7).



Figure 6 : Différence de composition de carcasse pour quatre races bovines à âges similaires. Adapté de Gotoh et al. (2009).



Figure 7 : Différence de répartition des TA pour quatre races bovines à âges similaires. Adapté de Gotoh et al. (2009).

Du fait des différences de répartition anatomique de pourcentage des dépôt adipeux dans la carcasse à âge comparable, les races bovines internationales ont été classées en 3 grandes types (Pethick, Barendse, Hocquette, Thompson et Wang, 2007):

- Les races précoces (*early-maturing*), qui déposent précocement les TA et donc qui, à âge similaire, ont plus de dépôt adipeux que les races tardives. Ces races sont les Japanese Black et Angus.
- Les races tardives (*late-maturing*) qui sont des races avec une croissance du dépôt adipeux plus tardive au regard de la croissance du tissu musculaire, dite race maigre. Ces races maigres sont par exemple les Belgian blue ou Blanc Bleu Beige, hypermusclé du fait de la mutation du gène myostatine et de l'absence de la protéine correspondante.
- Les races intermédiaires comprenant des races telles que les races Holstein, les Salers, Montbéliarde ou Rouge des prés (Cundiff et al., 1993) considérées comme tardives relativement aux races précoces et inversement par rapport aux races tardives.

De manière générale, les bovins des races continentales (type races tardives), ont entre 15 et 18% de lipides dans la carcasse et ont un poids plus élevé à l'âge adulte que les races anglaises Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines

(type races précoce), qui sont à plus de 25% de lipides dans la carcasse (Al-Jammas, 2017; Kempster, 1986). Les races à viande, souvent des races tardives, ont des proportions plus élevées de dépôts sous-cutanés que les races laitières (vaches de réformes utilisées ensuite pour la production de viande) qui ont des dépôts adipeux internes plus élevés. D'autre part, en terme de lipides intramusculaires (LIM), les races à viande continentales, donc tardives, ont moins de gras intramusculaire comparées aux races précoces, mais plus de gras intermusculaire (Tableau 4).

Gotoh et al., (2009), ont montré que la relation entre les dépôts de lipides dans la carcasses et le muscle était linéaire pour les races tardives et intermédiaire et quadratique pour les races précoces, suggérant que les facteurs de variations qui affectent le dépôt de lipides dans la carcasse affectent aussi le dépôt de lipides musculaires (Figure 8).



Fig. 6. Relationships between the IMF content in longissimus muscle and weights of subcutaneous, intestinal, omental, and perirenal adipose tissue in growing Japanese Black, double-muscled Belgian Blue, German Angus, and Holstein-Frieslan cattle. Regression equations were determined as follows: (a) $y = 3.0140 \times + 4.5692 \ (r^2 = 0.84)$, (b) $y = 1.3787 \times + 4.0667 \ (r^2 = 0.81)$, (c) $y = -0.0607 \ \pi^2 + 2.255 \ x + 0.9256 \ (r^2 = 0.84)$, (d) $y = -0.0422 \ \pi^2 + 2.1844 \ x + 0.4501 \ (r^2 = 0.85)$, (e) $y = 10.7460 \ x - 0.2810 \ (r^2 = 0.68)$, (f) $y = 4.4637 \ x + 0.04501 \ (r^2 = 0.62)$, (g) $y = 6.7826 \ x - 0.01600 \ (r^2 = 0.68)$, (h) $y = 9.5163 \ x - 0.4219 \ (r^2 = 0.78)$, (i) $y = 7.972 \ x - 1.9366 \ (r^2 = 0.74)$, (j) $y = 2.3311 \ x + 0.0200 \ (r^2 = 0.68)$, (h) $y = 4.3218 \ x - 0.2200 \ (r^2 = 0.89)$, (n) $y = 2.1586 \ x + 0.2372 \ (r^2 = 0.91)$, (o) $y = 3.7888 \ x - 0.2105 \ (r^2 = 0.91)$, (p) $y = 4.3705 \ x - 0.2476 \ (r^2 = 0.92)$, where y: weights of adipose tissues and x: IMF content in LM.

Figure 8 : Relation souvent linéaire entre le dépôt adipeux intramusculaire et le poids de la carcasse provenant de Gotoh et al. (2009). IMF : lipide intramusculaire ; weight : poids de carcasse.

Ainsi, les races précoces dont la croissance musculaire s'atténue rapidement avec l'âge (Angus, Noir Japonais, Holstein) déposent plus de lipides que les races tardives au potentiel de croissance musculaire plus élevé (Blanc Bleu Belge, Blond d'Aquitaine, Limousin) (Bonnet et al., 2007; Gotoh et al., 2009). Comparativement aux races tardives, les races précoces ont en général un nombre d'adipocytes plus élevé, des adipocytes plus gros et des activités

lipogéniques plus fortes (Bonnet et al., 2010). Il n'est donc pas toujours aisé de comparer des animaux de races différentes en fonction de l'âge. Une manière de comparer ces animaux est de calculer un âge physiologique ou un degré de maturité qui peut être estimé comme le poids à l'abattage divisé par le poids à l'âge adulte.

Tableau 4 : Pourcentage de Lipides intramusculaires (LIM) pour différentes races. n nombre d'animaux pour chaque race ; ()— Ecart type; a,b: Dans une ligne, les valeurs ne partageant pas un exposant commun sont significativement différentes pour la race. Adapté de Cafferky et al. (2019).

| | Races | | | | | | | |
|-----|-------|----------|----------|-----------|--------|-----------|-----------------|-------------|
| | Angus | Hereford | Limousin | Charolais | Salers | Simmental | Belgium Blue | Parthenaise |
| n | 64 | 13 | 296 | 168 | 41 | 79 | 77 | 15 |
| LIM | 2.78 | 2.16 | 2.13 | 2.05 | 2.41 | 2.13 | 1.7 | 1.12 |
| (%) | (0.2) | (0.5) | (0.1) | (0.1) | (0.2) | (0.3) | (0.2) | (0.4) |
| | b | ab | ab | ab | ab | ab | a | a |

1.4.3 Le sexe

Comparativement aux mâles, les femelles ont un développement plus précoce des TA bien qu'elles aient une croissance pondérale plus faible que ceux-ci et des teneurs en LIM supérieures. Inversement, la croissance des tissus musculaires est supérieure chez les mâles entiers. En effet, les taureaux ont été montrés comme ayant des muscles avec une plus grande vitesse de croissance notamment due à des activités de calpastatin et cystatin plus importantes que celles des bouvillons (Morgan, Wheeler, Koohmaraie, Crouse et Savell, 1993). D'autre part, les mâles castrés ont une vitesse de développement adipeux et une croissance pondérale intermédiaire entre celles des mâles entiers et des femelles, et ce pour une même durée de finition et avec la même alimentation (Steen et Kilpatrick, 1995). En outre la teneur en LIM des mâles castrés est souvent proche de celle mesurée dans le muscle de femelles et inférieure à celle des muscles de mâles entiers (Tableau 5). La variation de dépôt lipidique dans la carcasse et les muscle selon le sexe, résulte du rôle des hormones endogènes : notamment androgènes et œstrogènes qui régulent les enzymes de la lipogenèse. La testostérone promeut la différenciation des cellules mésenchymateuse en myocyte (en induisant l'expression de marqueurs myogéniques tels que MyoD) module les métabolismes musculaire (Picard, Gagaoua, Al Jammas et Bonnet, 2019) et inhibe la différenciation adipogénique en réprimant des marqueurs adipogéniques tels que PPAR γ 2 (Zhang et al., 2011). Inversement, la présence

Synthèse bibliographique - Productions et qualités des carcasses et viandes bovines

d'œstradiol chez les mâles castrés, résultant de la diminution de testostérone, augmente les activités des enzymes de la lipogenèse dans le TA sous-cutané (Prior, Smith, Schanbacher et Mersmann, 1983).

Tableau 5 : Pourcentage de lipide intramusculaire (LIM) selon les races et selon le sexe des animaux. ET : Ecart Type. LT : Longissimus Thoracis, LD : Longissimus Dorsi. LL : Longissimus Lumborum. KC : Korean cattle. Adapté de (Park, Beak, et al., 2018).

| Race | No. d'animaux | Sexe | LIM±ET (%) | Muscle |
|----------|---------------|-------------------|------------|--------|
| Wagyu | 6 | bouvillon | 34.3 ±1.7 | LT |
| Wagyu | 10 | vache | 37.8 ±1.00 | LT |
| KC | 41 | bouvillon | 13.3 ±4.40 | LD |
| KC | 10 | vache | 14.3 ±0.43 | LL |
| KC | 20 | bouvillon | 15.3 | LT |
| KC | 77 | bouvillon | 13.5 | LD |
| Angus | 10 | bouvillon | 14.7 | LL |
| Angus | 450 | bouvillon | 6.2 ±0.21 | LT |
| Angus | 9 | bouvillon | 7.1 | LD |
| Hereford | 10 | bouvillon | 8.3 | LL |
| Brahman | 7 | bouvillon | 3.1 | LD |
| Brahman | 10 | Taureaux et vache | 2.8 | LT |
| Brahman | 10 | Taureaux et vache | 1.9 | LD |
| Brahman | 50 | bouvillon | 3.4 | LD |

1.4.4 L'alimentation

Le métabolisme lipidique varie avec le bilan énergétique des animaux qui lui-même dépend de la nutrition et des besoins en énergie liés à l'entretien et à la production. Une augmentation des quantités ingérées favorise le dépôt de lipides dans la carcasse et le muscle, via une augmentation des activités lipogéniques concomitante avec une augmentation de la disponibilité en nutriments précurseurs de la lipogenèse. À l'inverse, lorsque le niveau d'alimentation ou d'énergie est faible, les activités lipogéniques sont réduites et les nutriments ne sont que faiblement stockés dans les adipocytes. La nature des nutriments apportés modifie également l'intensité de la lipogenèse. Par exemple, la disponibilité en substrats de la lipogenèse *de novo*, le glucose notamment, est augmentée par des rations riches en concentrés, apportant à la fois énergie et glucides, augmentent les activités des enzymes de la lipogenèse *de novo dans le muscle* (Schoonmaker et al., 2004). Jusqu'à récemment il était admis que le glucose était le précurseur majeur de la synthèse de lipides musculaires (S. B. Smith et Crouse,

1984), ce qui a été remis en question par des travaux récents montrant que l'acétate contribuait aussi à la biosynthèse des acides gras dans les tissus adipeux intramusculaires et sous-cutanés (Choi et al., 2014; S. C. Lee, Lee, Kim, Park et Han, 2000; Miller, Cross, Lunt et Smith, 1991; S. B. Smith et al., 2018)

Les adiposités corporelle et musculaire des bovins sont donc influencées par plusieurs facteurs (âge, race, sexe et l'alimentation) et dépendent d'évènements cellulaires et moléculaires de l'hypertrophie et de l'hyperplasie des tissus adipeux relativement aux muscles.

CHAPITRE 2 : LES METHODES OMIQUES

2 OBTENTION ET DISPONIBILITE DES DONNEES TRANSCRIPTOMIQUES ET PROTEOMIQUES CHEZ LE BOVIN

2.1 LES METHODES POUR L'OBTENTION DES DONNEES PROTEOMIQUES

Ces méthodes ont pour objet de quantifier l'abondance de toutes les protéines présentes dans un tissu ou une cellule. La quantification et l'identification des protéines peuvent se faire via 2 types d'analyses : la protéomique basée sur les gels (*gel-based proteomics*) ou la protéomique sans gel de séparation des protéines (*gel-free proteomics*).

2.1.1 Protéomique basée sur les gels

Après extraction, les protéines sont séparées en deux étapes successives. L'électrophorèse bidimensionnelle (2-DE), sépare les protéines selon le pH ou pH isoelectrique dans une dimension (étape d'électroisofocalisation) puis selon la masse molaire dans l'autre dimension (gel de polyacrylamide de type SDS-PAGE). Ainsi, les protéines sont séparées et constituent des points individuels ou spot, chaque spot correspondant le plus souvent à une protéine majoritaire. Une analyse d'image est réalisée après une coloration afin de quantifier l'abondance de chaque point (protéine). En général, il y a comparaison d'abondance de points entre gels pour déterminer si certaines protéines sont identifiées comme différentiellement abondantes entre les traitements étudiés. Dans le cas où des protéines sont identifiées comme différentiellement abondantes, le spot est découpé sur le gel puis traité avec des protéases afin d'obtenir des peptides spécifiques de la protéine. Ces peptides sont ensuite identifiés par spectrométrie de masse (MS).

Il existe différents types de méthodes basées sur des séparations sur gel qui différent selon les conditions de séparation des protéines (pH, poids moléculaire...) et les méthodes de révélations des protéines (fluorescence, colorimétrie...).

2.1.2 Protéomique sans gel

Dans le cas de la protéomique sans gel, les protéines présentes après extraction sont digérées par des enzymes protéolytiques telle que la trypsine conduisant à la synthèse de petits

peptides. C'est cet ensemble de peptides qui est ensuite analysé par spectromètre de masse. Parfois des étapes intermédiaires sont incluses telles qu'une concentration des protéines (sur gel par exemple) ou un marquage chimique ou isotopique. Par la suite, les peptides sont séparés et identifiés en utilisant un système de chromatographie liquide couplé à un spectromètre de masse en tandem (LC-MS/MS). Les peptides sont représentés par des pics formant des spectres expérimentaux. L'étude de ces spectres permet d'identifier une protéine en les comparant à ceux répertoriés dans une banque de données pour les protéines connues dans une espèce. La méthode la plus couramment utilisée est appelée shotgun.

2.2 LES METHODES POUR L'OBTENTION DE DONNEES TRANSCRIPTOMIQUES

2.2.1 Puce à ADN

Cette méthode a pour objet de quantifier l'abondance de tous les ARN présents dans un tissu ou une cellule. Le principe des puces à ADN ou biopuces (*micro-array* en anglais) repose sur la propriété de l'ADN et de l'ARN à s'hybrider (reconnaissance et interaction de deux séquences complémentaires d'ADN ou d'ARN). Les puces sont des supports solides (des plaques de verre par exemple) sur lesquels sont déposées des sondes. Ces sondes peuvent être des assemblages d'oligonucléotides ou des ADNc. Par souci de simplification, nous ne détaillerons pas les étapes permettant d'obtenir une puce de bonne qualité, mais seulement le principe de fonctionnement. Sur les puces, les sondes sont simples brins et peuvent s'hybrider avec les séquences cibles des ADNc issus de la transcription inverse des ARNm par exemple. Dans le cas des ARNm, les ADNc sont marqués avec des fluorophores, les plus courants étant Cy3 et Cy5 pour respectivement Cyanine 3 et Cyanine 5 (Figure 9). La fluorescence de chaque fluorophore est ensuite quantifiée à l'aide d'un scanner (après lavage de la puce pour enlever les séquences non hybridées), et est proportionnelle à l'abondance de l'ADNc de l'ARNm à quantifier.



Figure 9: Schéma de fonctionnement des puces à ADN provenant de Duggan, Bittner, Chen, Meltzer et Trent (1999).

Il existe des puces à 1 ou 2 canaux de lecture, selon le nombre de fluorophores utilisés pour l'hybridation. Les puces à un fluorophore, appelées puces monochromatiques, ont un canal de lecture de la fluorescence et analysent un échantillon par condition expérimentale. Les puces à 2 fluorophores, ont 2 canaux de lecture et quantifient simultanément l'abondance d'ARNm dans 2 échantillons divergents et utilisent le principe d'hybridation compétitive. Le traitement des données commence par la normalisation de la fluorescence et les méthodes dépendent du type de puce. Les plus courantes appliquent une mise à l'échelle et un centrage médian (« scaling ») ou une normalisation par la méthode Loess (Guerra et Goldstein, 2016).

2.2.2 Séquençage ARN

Cette méthode a pour objet de quantifier l'abondance de tous les ARN présents dans un tissu ou une cellule. Le séquençage ARN, ou RNA-seq est également appelé séquençage aléatoire du transcriptome entier (*whole transcriptome shotgun sequencing*). Il utilise le

séquençage haut débit (*next-generation sequencing* ou NGS) afin d'identifier l'ARN issu de la transcription du génome dans un échantillon biologique. Selon les ARN extraits et préparés pour le séquençage, cette technologie identifie des ARNm, les microARN (appelé également miRNA), les ARN de transfert (ARNt) ou encore les ARN ribosomaux (ARNr).

La méthodologie est adaptée au type d'ARN ciblé. Dans le cas des ARNm, les ARN sont isolés et sélectionnés en visant leurs queues polyadénylées (polyA). Les séquences polyA se fixent sur des séquences oligonucléotides polyThymine (polyT) elles-mêmes fixées à des billes magnétiques (Figure 10). Les ARNm sont hydrolysés en fragments de 200 à 300 nucléotides (afin de réduire la formation de structures secondaires et corriger le biais 5' de la transcription inverse) et rétro transcrits en ADNc via l'utilisation d'amorces aléatoires très courtes et non spécifiques. Les fragments d'ADNc sont ensuite fragmentés à une longueur déterminée pour le système de séquençage (Illumina étant la plus utilisée, pyroséquençage ou encore séquençage SOLiD). Dans le cas de l'utilisation de la technologie Illumina, des fragments d'ADNc auxquels ont été fixés des adaptateurs sont dénaturés en ADN simple brin. Ces simples brins se fixent, via les adaptateurs à la surface du « flowcell » sur lesquels ils vont subir une étape de « PCR bridge » consistant en la création d'un ADN double brin à partir de la matrice simple brin. Il y a à nouveau dénaturation de l'ADN, et par répétition création de groupes (ou cluster) denses où les fragments sont amplifiés (Figure 11).



Figure 10 : Schéma du fonctionnement de la fixation des ARNm lors d'un séquençage ARNm (RNA-seq)



Figure 11 : Schéma de récupération des billes et fonctionnement de la reverse transcriptase lors d'un séquençage d'ARN

Le séquençage se fait ensuite par cycle avec ajout des 4 bases marquées, des amorces et de l'ADN polymérase : les amorces se fixent et la polymérisation commence en utilisant les bases marquées. Après l'ajout de chaque base, il y a excitation par un laser et la fluorescence permet de lire la 1^{re} base ajoutée dans chaque cluster. Le cycle continue en lisant les bases une par une dans chaque cluster. La séquence cible correspond au brin complémentaire des bases lues (par exemple pour une lecture ACCTG, la séquence sera TGGAC).

L'analyse informatique des données a pour objectif d'aligner les séquences de fragment d'ADNc, appelé également « read », sur le génome de l'espèce de référence (après un contrôle de qualité du séquençage). Les séquences à aligner étant souvent courtes, plusieurs séquences s'alignent au même endroit. Le nombre moyen de séquences par position est alors mesuré et est appelé la profondeur de séquençage. Pour chaque région d'intérêt, le nombre de reads est comptabilisé et permet ainsi la création de tableaux de comptage, qu'il faudra ensuite normaliser pour analyser et ainsi quantifier l'abondance de l'ARNm.

Le RNAseq est une méthodologie assez récente qui n'a été vraiment développée et utilisée qu'à partir de 2007 (Emrich, Barbazuk, Li et Schnable, 2007) comparée aux puces à ADN qui sont utilisées depuis 1987 pour l'analyse d'un profil d'expression (Lashkari et al., 1997). Bien que les techniques de RNAseq soient de plus en plus utilisées, les données qu'elles génèrent sont souvent non disponibles sur Internet. Les données provenant de puces à ADN restent pour l'instant majoritaires, notamment pour les animaux de rente, moins étudiés que la souris ou l'Homme. Cependant, les données générées par les puces à ADN sont diverses, de par les méthodes de marquage et lecture qui impactent aussi les méthodes de normalisation et la comparaison des résultats.

3 DONNEES PROTEOMIQUES APPLIQUEES A LA COMPREHENSION DE L'ADIPOSITE CORPORELLE ET TISSULAIRE DES BOVINS

Les données protéomiques relatives à l'adiposité corporelle ou musculaire chez le bovin sont exclusivement disponibles dans des publications, avec des abondances de protéines qui ne sont pas toujours rapportées. Aucune donnée brute facilement réutilisable n'a été déposée dans des bases de données dédiées comme proteomeXchange. La date de fin de recensement des données protéomiques est fin juillet 2019.

Cinq études rapportent des variations de l'abondance de protéines selon l'adiposité musculaire chez des bovins de races essentiellement précoces (Keady et al., 2013; Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Q. Zhang et al., 2010). Trois articles complètent ces données et sont cependant plus difficiles à interpréter. En effet ces études ont été réalisées à partir de muscle de *bos taurus* élevés au États-Unis (la race n'est pas précisée et il s'agit probablement d'une race précoce (Thornton, Chapalamadugu, Eldredge et Murdoch, 2017)), de *bos indicus* (Poleti et al., 2018) et d'Angus comparés à des *bos indicus* (De Souza Rodrigues et al., 2017)(Tableau 6).

Ces études ont mis en évidence pour des groupes de bovins divergents par la teneur en lipides du muscle, des protéines différentiellement abondantes et impliquées dans : les types contractiles lent et rapide du muscle (De Souza Rodrigues et al., 2017; Keady et al., 2013; Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Thornton et al., 2017; Q. Zhang et al., 2013; Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Thornton et al., 2017; Keady et al., 2013; Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Thornton et al., 2017) et oxydatif (Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Q. Zhang et al., 2010). D'autres protéines plus rarement identifiées interviennent dans la synthèses des acides gras (Keady et al., 2013), la signalisation par des protéines favorisant la mitose et la division cellulaire (Mao et al., 2016) et les signalisations ou régulations liées aux protéines de choc thermique (heat shock proteins ; De Souza Rodrigues et al., 2017; Y.-Y. Zhang et al., 2011). Cependant, toutes ces études ne sont pas

toujours d'accord sur le sens de variation de l'abondance des protéines identifiées selon la teneur en lipides.

Pour ce qui concerne les données protéomiques relatives à l'adiposité de la carcasse et en particulier à la croissance des tissus adipeux corporels, 4 études ont été réalisées dont trois avec la méthode 2 DE (Taga et al., 2012; Q. Zhang et al., 2010; Zhao, Basu, Dodson, Basarb et Guan, 2010) et une avec la méthode shotgun (Romao, Jin, He, Mcallister et Guan, 2013). Une première étude révèle les protéines différentiellement abondantes selon l'âge fœtal, pour comprendre les mécanismes précoces de la croissance des tissus adipeux. Lors des phases précoces de croissance fœtale, des différences d'abondance des protéines TCPI, FKBP4 et HSPD1, réguleraient la prolifération des précurseurs d'adipocytes (Taga et al., 2012). Lors des phases plus tardives et après 180 jours post-conception, il y a une augmentation de l'abondance des protéines induites par la voie PPARy en lien avec la différenciation terminale des adipocytes et l'augmentation du dépôt de lipides du fait de l'équilibre lipogenèse et lipolyse (Taga et al., 2012). Chez des bouvillons en croissance de race British-continental, l'induction des dépôts adipeux par un régime riche en lipides, modifie l'abondance de 39 et 15 protéines dans les tissus adipeux sous cutané et périrénal (Romao et al., 2013). L'analyse fonctionnelle a indiqué que les voies majeures associées aux protéines différentielles sont les métabolismes énergétique et lipidique. Parmi les protéines liées au métabolisme lipidique, il y a SCD, APOC3, ANXA6, PLIN1, MDH1, et ITGA6 (Romao et al., 2013). Parmi les protéines originales, la protéine ITGA6 (integrin alpha 6) pourrait induire l'arrêt de croissance des préadipocytes et favoriser la différenciation adipocytaire. Cette protéine pourrait également contribuer à la différenciation de nouveau précurseurs adipeux dans des situations de fort dépôt lipidique (Romao et al., 2013). Enfin, cette étude met aussi en évidence, que les abondances des protéines sont plus divergentes selon le site anatomique des tissus adipeux (240 protéines différentielles) que selon le régime alimentaire. Lorsque des différences d'épaisseur de gras sous-cutané ont été comparées chez des bouvillons Charolais × Red Angus ou Hereford × Angus, 36 protéines avec des abondances différentielles ont été identifiées dans le tissu adipeux sous-cutané (Zhao et al., 2010). Les protéines annexine 1 (ANXA1) et α-soluble NSF attachment protein (α-SNAP) ont été confirmées comme différentiellement abondantes selon l'épaisseur de gras sous-cutané, même si d'autres protéines ont été identifiées et sont impliquées dans les métabolismes glycolytique et énergétique telles que G3PDH, MDH1, IDH1 (Zhao et al., 2010). Enfin une dernière étude réalisée en 2DE sur des préadipocytes en différenciation, compare le protéome de cellules adipeuses préparées à partir de 3 sites anatomiques : omental, sous cutané et intramusculaire (Rajesh et al., 2010). L'abondance de 65 protéines est modulée quelle que soit l'origine anatomique des préadipocytes. Parmi celles-ci, l'auteur met en avant les protéines dont l'abondance augmente avec le dépôt de lipides, telle que l'abondance de UQCRC1, SOD, GAPDH, SULT1A1, CPT2 et HSPB1 impliquées entre autres dans les métabolismes oxydatif ou glycolytique. Les abondances de ces protéines sont plus élevées dans le tissu adipeux omental que dans les tissus adipeux sous-cutané et intramusculaire, en accord avec leur moindre activité métabolique. Cette dernière étude *in vitro* n'a pas été prise en compte dans le Tableau 6, rapportant le nombre de travaux disponibles par facteurs de variation et tissus.

Tableau 6: Nombre de publications protéomiques et de publications ou données transcriptomiques disponibles par tissu et facteur de variation de la croissance des tissus adipeux et des muscles chez le bovin. TA : tissu adipeux, TM : tissu musculaire. Pour les données transcriptomiques, le premier chiffre correspond au nombre d'étude et le deuxième correspond au nombre de jeux de données déposés dans les bases publiques sous forme de GSE. Un GSE peut contenir plusieurs GSM et donc des abondances de gènes selon différents traitements.

| Types de jeu de données | Protéomiques | | Puces à ADN | | RNAseq | |
|-------------------------|--------------|----|-------------|-----|--------|-----|
| | TA | TM | TA | TM | TA | ТМ |
| Age | 1 | 2 | - | 4+4 | - | 1+0 |
| Race | 1 | 4 | - | 3+4 | 1+0 | - |
| Sexe | 1 | 1 | - | 1+2 | - | - |
| Alimentation | 1 | 1 | 2+2 | 3+2 | 0+1 | 3+1 |
| Total | 4 | 8 | 4 | 23 | 2 | 5 |

4 DONNEES TRANSCRIPTOMIQUES POUR LA COMPREHENSION DE

L'ADIPOSITE CORPORELLE OU TISSULAIRE DES BOVINS

Les données transcriptomiques relatives à l'adiposité corporelle ou musculaire chez le bovin sont disponibles dans des publications, et dans les bases de données publiques dédiées. Dans les bases de données, les données disponibles sont brutes ou analysées par les auteurs, exprimées sous forme d'abondance pour une condition d'analyse, ou sous forme de ratio entre deux traitements. La date de fin de recensement des données transcriptomiques est fin juillet 2019.

Parmi les données disponibles, 28 jeux de données (23 issus de puces à ADN et 5 des données RNAseq) sont relatifs à la croissance ou la composition de différents muscles

(Semitendinosus muscle, Longissimus muscle, Longissimus lumborum, Longissimus dorsi ou Longissimus thoracis) et 6 (4 puces ADN et 2 RNAseq) sont relatifs à des variations de la masse des tissus adipeux corporels (tissus adipeux sous-cutané ou viscéraux). Ces jeux de données peuvent être catégorisés selon les facteurs de variations majeurs de l'adiposité corporelle ou musculaire (cf Tableau 6). Des variations de l'adiposité corporelle ou musculaire sont dues à l'âge dans 9 jeux de données, la race dans 8 jeux de données, le sexe dans 3 jeux de données et l'alimentation dans 14 jeux de données. A ce jour et pour les tissus adipeux et musculaire, 7 jeux de données (5 muscle et 2 TA) ont été produits par RNAseq avec comme facteurs de variation de l'adiposité corporelle ou musculaire l'âge pour une étude « muscle » (He et Liu, 2013), la race pour une étude « tissu adipeux » (Jin et al., 2015), et la nutrition pour une étude « tissu adipeux » et quatre études « muscle » (Baldwin, Li, Li, Thomson et Bequette, 2012; Carrillo et al., 2016; English et al., 2018; Keel et al., 2018; Keogh et al., 2016). L'essentiel des données a été obtenu avec des puces à ADN, et essentiellement sur le muscle et ont fait l'objet de 11 publications, certains jeux de données publiés sous forme de GSE n'ont pas donné lieu à une publication.

Parmi les publications transcriptomiques et RNAseq muscles, cinq études rapportent le transcriptome (2 de données de puces à ADN et 3 de données RNAseq) du muscle de bovins divergents par la teneur en lipides intramusculaires (LIM). Dans ces publications, les résultats concernent essentiellement les voies métaboliques ou les processus cellulaires et moléculaires identifiés grâce à des annotations d'ontologie de gène (GO). Il y a aussi quelques dizaines de gènes mis en avant par les auteurs, ce qui reste une vision parcellaire des données produites puisqu'il y a souvent plusieurs centaines de gènes identifiés comme différentiellement exprimés. Parmi les voies métaboliques ou les processus cellulaires et moléculaires souvent identifiés, on retrouve :

- La phosphorylation oxydative en lien avec le métabolisme glycolytique chez des bovins Holtein-Fresian, Brahman (Lim et al., 2014) ou Hanwoo (Seung-hwan Lee et al., 2010)
- La voie de signalisation de PPARγ et les voies du métabolisme des lipides chez des bovins de races Holtein-Fresian, Brahman (Lim et al., 2014), Hanwoo (Seung-hwan Lee et al., 2010), Charolais x Holstein-Fresian, (Komolka et al., 2016), Angus x Simmental similaires par l'épaisseur de gras dorsal mais différents par les LIM (Clark et al., 2011), Nellore (*bos indicus*) (Cesar et al., 2015), et Angus (D. Chen, Li, Du, Wu et Cao, 2015)

- Le métabolisme des glucides: chez des bovins de races Charolais x Holstein-Fresian (Komolka et al., 2016), notamment des gènes liés à la glycolyse (via la pyruvate kinase) et la voie des pentose phosphate (via la G6PD), mais aussi chez des Angus (D. Chen et al., 2015).
- Des transports moléculaires : chez des bovins de races Charolais x Holstein-Fresian (Komolka et al., 2016) et Angus (D. Chen et al., 2015).
- La croissance musculaire, la contraction musculaire et des modifications de la matrice extracellulaire: chez des bovins de race Angus-Simmental similaires par l'épaisseur de gras dorsal mais différents par les LIM (Clark et al., 2011)

A partir des annotations GO données dans les publications, il est souvent difficile de faire une synthèse précise des connaissances apportées, et des termes généralistes comme « transport moléculaire » sont souvent utilisés. Une autre façon de synthétiser l'information est de comparer les gènes différentiellement abondants listés dans la publication ou dans les informations supplémentaires associées aux publications (Tableau 7).

| | (Clark et al., 2011) | (Komolka et al., 2016) | (Cesar et al., 2015) | (Lee et al., 2010) | (Chen et al., 2015) |
|---|-------------------------------------|--|---|--|--|
| Méthodes | RNAseq | Puce à ADN | RNAseq | Puce à ADN | RNAseq |
| Gènes surexprimés avec les LIM | MYH3, HOXD10, MXRA8, CASQ2 | ACLY, AGRN, CIDEC, CILP, FABP4, ICMT, INSIG1, MGST1, MRAS, SCD, THRSP | (top 10 des genes DE) NCF4, ZFPM1, BCL2L12, GLIPR1, ZPF90, LRRC55, ATP6V1G2, MPZL1 | SH3YL1, ADAMTS4, TBC1D7, SQLE, CYP51A, TUBD1 | (top 10 des gènes DE) KLF9, CLCN4, BT.24356, MLL, MYNN, ATPAF1, JHDM1D, FKBP3, RAI14, RMND5A |
| Gènes sous- exprimés avec les LIM | NPNT, MRC1, DNER, CYPB4 | ART3, ATXN7L1, BACH2, C8orf22, CDC40, CETN3, CIR, FRAT2, GNAS, GPD2, HIST1H1E, IL4R, MGC13057, PROX1, RAD50, SESN1, SLC43A2, SMARCAD1, TAP2, TBXAS1, TRA | (top 10 des genes DE) CSTF1, ANKRD26, NCR3, MYRF, ABHD8, LRRC16B, HOXC5, PPAPDC2, ANKRD26 | SH3KBP1, TXNDC9, CDClk1, RARs, UTP15, BOLA- DMA, THOP1 | (top 10 des gènes DE avec le plus grand FC) CNN1, BT.102070, SBK1, TGFB3, NKG7, IMPA2, FBP1, CADM4, L1CAM, TMEM35) |

Tableau 7 : Gènes différentiellement exprimés entre des animaux divergents par la teneur en lipides intramusculaires (LIM) dans 5 études.

Même si les voies métaboliques identifiées semblent communes à différentes études et races, lorsque l'on compare les listes de gènes différentiellement abondants, très peu sont communs entre études, probablement en raison de nombres de gènes listés différents (Tableau 7 ; Figure 12). Pour les cinq études comparées, aucun gène n'est commun entre les cinq listes, 3 gènes (AGRN, SCD et THRSP) sont surexprimés dans deux listes (D. Chen et al., 2015; Komolka et al., 2016) et un autre gène (ATP6V1G2) est surexprimé dans deux autres listes (Cesar et al., 2016; D. Chen et al., 2015). Cependant certains gènes sont surexprimés dans une étude et sous-exprimés dans une autre étude : CASQ2 (surexprimé (Clark et al., 2011) et sous-exprimé (D. Chen et al., 2015)), SQLE (surexprimé (Seung-hwan Lee et al., 2010) et sous-exprimé (D. Chen et al., 2015)) et SNORD15 à la fois surexprimé et sous-exprimé dans la même étude (D. Chen et al., 2015).



Figure 12: Diagramme de Venn des gènes différentiellement exprimés décrits dans le Tableau 7

Pour conclure, il est difficile de synthétiser les connaissances apportées par les données transcriptomiques décrites dans les publications, en raison de la diversité des facteurs de variations (croissance musculaire *versus* teneur en LIM), de la diversité des modes de présentations des résultats qui synthétisent souvent l'information sous forme d'annotation GO ou de listes de quelques gènes. Enfin, il n'y a que peu de consensus entre les études, probablement en raison des facteurs de variation souvent spécifiques à une étude (une race dans une condition d'élevage et à un âge donné) ou du faible nombre de réplicats biologiques, lié au coût des méthodes « omiques ».

Il est donc nécessaire de trouver un moyen d'extraire, agréger, et re-analyser les résultats de plusieurs études pour vérifier les connaissances acquises et en augmenter la fiabilité. En outre, un grand nombre de données disponibles sous forme de GSE n'ont pas été toutes publiées et n'ont pas encore livrées de connaissance. Une des méthodes pour avoir une vision globale de l'ensemble de ces travaux est de combiner statistiquement les données protéomique ou transcriptomique par la méta-analyse.

Synthèse bibliographique – Méthodes omiques

I

CHAPITRE 3 : LA META-ANALYSE

5 INTRODUCTION A LA META-ANALYSE

La méta-analyse est la synthèse statistique systématique de résultats provenant de plusieurs études indépendantes, conduites pour répondre à une question biologique donnée (Borenstein, Hedges, Higgins et Rothstein, 2009; Dawson, Pihlstrom et Blanchette, 2016; Ramasamy, Mondry, Holmes et Altman, 2008; Tseng, Ghosh et Feingold, 2012; Williams, 2012). Le terme méta-analyse est souvent utilisé à deux fins : décrire le processus complet de production de connaissances par l'analyse de résultats d'études indépendantes, ou décrire une méthode statistique pour le traitement de données issues de plusieurs études. La description du processus, est également nommée revue systématique de la littérature (Ramasamy et al., 2008; Tseng et al., 2012). La combinaison de résultats par la méta-analyse augmente le nombre de cas étudiés et ainsi la précision des découvertes. Cette approche valide ou non les conclusions des études indépendantes, assurant un contrôle des connaissances existantes (Tseng et al., 2012).

L'histoire de la méta-analyse a fait l'objet d'une synthèse réalisée par Gurevitch et al., (Gurevitch, Koricheva, Nakagawa et Stewart, 2018). Brièvement la première tentative d'agrégation d'informations provenant de plusieurs études a été réalisée en 1904 par K. Pearson (Pearson, 1904) afin de vérifier l'efficacité de vaccination contre la typhoïde de soldats. Différentes méthodes de méta-analyses ont commencé à être décrites à partir de 1938 (Yates et Cochran, 2006) et ont vraiment été formalisées par Cochrane en 1954 (Cochrane, 1954). Depuis, l'utilisation de la méta-analyse est en constante augmentation, comme en atteste le nombre grandissant par année de publications référencées dans PubMed et rapportant les termes « méta-analyse » ou « revue systématique » (Figure 13).





Figure 13: Nombre d'articles par année faisant référence à "meta-analysis" ou "systematic review" dans PubMed (1948-2019). Recherche faite le 11/09/2019.

Les avantages de la méta-analyse selon Ramasamy *et al.* (2008) sont nombreux, parmi lesquels :

- Réutiliser des données publiques
- Augmenter la fiabilité des résultats des études et généraliser ou infirmer des résultats individuels
- Utiliser une stratégie de recherche relativement peu chère, puisqu'elle utilise des données déjà existantes

Parmi ces avantages, la réutilisation des données publiques est un concept qui prend de l'importance. Il a d'abord été évoqué en 2013 par Rung et Brazma (Rung et Brazma, 2013) concernant les données transcriptomiques publiques et notamment les données de puces à ADN (en prédisant une future utilisation des données RNA-seq, une fois les protocoles stabilisés et les coûts diminués). La réutilisation des données a également été traitée par Vaudel et al. (2016) pour les données publiques de protéomique. Quels que soient les domaines d'application, ces publications mettent en avant que la réutilisation des données publiques augmente la compréhension des phénotypes biologiques sans nouvelle expérimentation et affine des hypothèses de recherche. Ces deux finalités sont en accord avec les règles de réduction et de raffinement des principes éthiques des 3R en expérimentation animale. Les principes éthiques des 3R sont de Réduite le nombre d'animaux utilisés dans les expérimentations, Raffiner la méthodologie utilisée (impliquant les critères d'interruption) et enfin Remplacer les modèles animaux si cela est possible (Russell et Burch, s. d.).

Cette thèse se situe dans un contexte de traitement de données « omiques » et plus particulièrement de données issues de puces à ADN. Les descriptions faites dans la suite du chapitre sur la méta-analyse seront donc spécifiques à la méta-analyse de données omiques. L'objectif de ce chapitre sur la méta-analyse est de donner une description générale des données utilisées pour les méta-analyses, puis de préciser un plan d'analyse, des méthodes et outils classiquement utilisés pour la méta-analyse de données transcriptomiques. En aucun cas, il ne s'agit d'une revue des méthodes mathématiques et statistiques utilisées en méta-analyse.

6 LES DONNEES POUR LA META-ANALYSE

Deux types de données sont utilisées en méta-analyse : les données qualitatives ou quantitatives (Smith et al., 2016). Ces données sont soit des données individuelles (*individual participant data*, abrégé en IPD en anglais et DIP en français) ou des données agrégées ou résumées (*aggregate data*, abrégé en AD, DA en français) correspondant souvent à des données observées sur un ensemble d'individus ou d'échantillons (Debray et al., 2015; Maison, 2010) (Figure 14). Les données utilisées dans les méta-analyses transcriptomiques sont majoritairement de type quantitative (l'analyse portant sur l'abondance des gènes).



Figure 14 : Diagramme des types de données et de méthodes classiquement utilisées en méta-analyse

individuelles correspondent Les données aux données de chacune des observations/mesures réalisées dans l'étude. La méta-analyse de données individuelles apporte une grande robustesse en distinguant l'hétérogénéité provenant d'un individu et celle provenant de l'étude entière. Cependant, il est souvent difficile d'obtenir l'ensemble des données individuelles de toutes les études pour une méta-analyse. De plus, la méta-analyse de données brutes peut être compliquée par une variabilité des expérimentations entre les études, l'introduction de covariables dans des études, ou encore le manque de descriptions pertinentes d'observations, paramètres ou traitements pour chaque étude à inclure dans la méta-analyse (Debray et al., 2015; Maison, 2010).

Les méta-analyses des données agrégées sont les plus courantes et sont également appelées méta-analyses de résultats basées sur la littérature. Les données agrégées correspondent à des informations globales pour des observations/mesures. Il s'agit par exemple de données agrégées sous forme de moyenne, médiane, de fréquence d'inclusion de données binaires ou de données de survie (Maison, 2010; Riley, Simmonds et Look, 2007)

Il est possible de combiner et de traiter des données individuelles et des données agrégées dans une méta-analyse, ce qui conduit à 4 types de méta-analyse (Debray et al., 2015; Riley et al., 2007):

- Une méta-analyse de données agrégées à partir de données individuelles : les données individuelles sont dans un premier temps transformées en données agrégées
- Une méta-analyse des données individuelles reconstruites : il est possible de reconstruire des données individuelles à partir de données agrégées/résumées pour des données binaires dans des tableaux 2x2 (Riley et al., 2007), des données ordinales (basées sur le nombre de réponses dans chaque condition) et des données de survie. Cependant, il est à noter que les données individuelles reconstruites n'ont pas de description pour chaque observation et cela peut biaiser ou nuire à l'interprétation des différences entre études.
- Méta-analyse par régression liée à la hiérarchie (*Hierarchical-related regression*) ou modèle des paramètres partagés (*shared parameters models*). Il s'agit d'une approche bayésienne qui calcule la probabilité de chaque type de données (données individuelles ou données agrégées) pour modéliser les réponses individuelles. Les probabilités peuvent être analysées en utilisant par exemple, les distributions telles que Bernoulli (pour les données individuelles) ou binomiales/normales (pour les données agrégées).

Ainsi, chaque étude participe à l'estimation de l'effet d'un traitement (Debray et al., 2015; Riley et al., 2007).

Une méta-analyse par modèle multi-niveaux (multilevel model) : un modèle dans lequel le plus haut niveau correspond à l'étude et le plus bas niveau correspond à une observation/mesure individuelle. Les données agrégées et individuelles sont incorporées par l'inclusion d'une variable factice (par exemple « données agrégées » et « données individuelles » pour préciser la source des données) dans le modèle. Le modèle distingue les résultats obtenus par les données agrégées et les données individuelles (Goldstein, Yang, Omar, Turner et Thompson, 2000; Riley et al., 2007).

Cependant, l'incorporation de données agrégées et de données individuelles dans une même méta-analyse n'augmente pas nécessairement la précision de la quantification de l'effet d'un traitement. Il est souvent préférable de conduire une méta-analyse uniquement sur des données individuelles puis de comparer les résultats obtenus à ceux produits par la méta-analyse sur des données individuelles et agrégées (Riley et al., 2007). Cette comparaison évalue l'impact des données agrégées sur la généralisation de la réponse, et l'hétérogénéité entre études ou l'hétérogénéité entre les observations individuelles (Debray et al., 2015; Riley et al., 2007).

7 PLAN D'ANALYSES POUR LA META-ANALYSE DES DONNEES TRANSCRIPTOMIQUES

Selon Ramasamy (2008) et Rung & Brazma (2013) il y a généralement 7 étapes critiques pour une méta-analyse sur des données de puces à ADN (cf Tableau 8) :

- 1. Identifier les jeux de données adaptés
- 2. Extraire les données des études
- 3. Préparer les jeux de données provenant de différentes plateformes
- 4. Annoter les jeux de données
- 5. Concaténer et mettre à jour les identifiants des gènes (ID)
- 6. Choisir une méthode de méta-analyse
- 7. Analyser, présenter et interpréter les résultats
Ces étapes semblent consensuelles pour l'intégration de données transcriptomiques en physiologie animale. Les étapes majeures (Figure 15 et Figure 16) ont en effet été décrites dans les rares méta-analyses de données transcriptomiques publiques chez le bovin pour identifier des indicateurs de la maturité des follicules ovariens (Khan et al., 2016), les signatures moléculaires génériques à la lactation (Farhadian, Rafat, Hasanpur, Ebrahimi, & Ebrahimie, 2018), et celles spécifiques de muscles sains versus pathologiques (Baron et al., (2011). Ces étapes majeures ont également été identifiées par Tseng et al. (2012), qui ont fait une métaanalyse pour caractériser les méthodologies utilisées dans 333 publications de méta-analyses de données de puces à ADN. Par rapport aux étapes décrites par Ramasamy et al. (2008) (Tableau 8), Tseng et al. (2012) s'intéressent particulièrement aux étapes 6 (choisir une méthode de méta-analyse) et 7 (analyser, présenter et interpréter les résultats). Ils mettent ainsi en évidence que la détection des voies métaboliques, associées à la mise en évidence de gènes différentiellement exprimés (DEG) est plus cohérente et plus reproductible que la seule analyse des DEGs. Les multiples actions à conduire par étape sont résumées dans le Tableau 8. Dans la suite de cette partie, j'illustrerai certaines étapes et actions jugées critiques par Tseng et al., (2012), Ramasamy et al., (2008) et Rung & Brazma (2013).

Synthèse bibliographique - Méta-analyse



Figure 15: Description du plan d'analyse de la méta-analyse de Farhadian et al 2018



Figure 16: Description du plan d'analyse proposé par Rung et Brazma 2013

Tableau 8 : Les différentes étapes d'une méta-analyse conduite sur des études de puces à ADN (adapté de Ramasamy et al., (2008))

- 1 Identifier les jeux de données relatifs à une question de recherche
 - Formuler les objectifs de l'étude et un protocole de sélection des études
 - Définir des mots clés pour une recherche bibliographique sur les sites Internet appropriés
 - Définir des critères d'inclusion et d'exclusion et approprié
 - Chercher des dépôts de données publiques de puces à ADN
 - Analyser la section des références des publications sélectionnées pour trouver d'autres études pertinentes
 - Analyser les études sélectionnées selon les critères de sélection ou d'exclusion
- 2 Extraire les données des études
 - Analyser les études/publications et URL relatives aux études pour identifier les caractéristiques des données
 - Si les données sont utilisées dans plusieurs publications et sont redondantes, extraire les données de la publication la plus complète
 - Augmenter le nombre de données en combinant les jeux de données d'entrainement et de validation provenant d'une même étude
- 3 Préparer les jeux de données provenant de différentes plateformes
 - Identifier et enlever les données de mauvaise qualité (en calculant les scores de qualité incluant la concordance des spots, la corrélation entre les 2 canaux de couleur et une mesure de robuste du bruit pour chaque hybridation (Buness, Huber, Steiner, Sültmann et Poustka, 2005))
 - Prétraiter les données pour enlever les différences systématiques entre les données d'expression, provenant de différentes plateformes par exemple en utilisant un algorithme de normalisation stabilisant la variance.
 - Vérifier les effets de groupes par tableau/jeu de données, spécifiquement dans les gros jeux de données
 - Enlever par filtrage les sondes avec de mauvaise qualité de détection dans le tableau de résultats (optionnel), en utilisant des indicateurs de qualité de la sonde produite par les systèmes de puces à ADN
 - Agréger tous les réplicas techniques
 - Vérifier que les valeurs d'expression issues de différentes plateformes sont compatibles
- 4 Annoter les jeux de données et les catégoriser
 - Identifier (a) la séquence de la sonde, (b) l'information d'annotation de la sonde la plus spécifique à la séquence, (c) les informations relatives à la description des échantillons biologiques pour catégoriser les données.
- 5 Résoudre les multiples relations entre les sondes et les identifiants de gènes
 - Mettre à jour les identifiants de gènes à partir des séquences des sondes
 - Enlever les sondes qui ne se superposent à aucun identifiant de gène
 - Pour chaque identifiant de gène dans une étude, calculer l'estimation de la spécificité à l'étude en calculant si possible la taille d'effet et la variance associée à la taille d'effet pour chaque gène.
 - Si une sonde se superpose à de multiples identifiants de gène dans une étude, « étendre » celle-ci en la remplaçant avec un nouvel identifiant de gène avec le même profil d'expression
 - Pour tous les identifiants de gènes avec de multiples enregistrements dans une étude, les « résumer » soit en sélectionnant un seul des enregistrements soit en les agrégeant
- 6 Choisir une méthode de méta-analyse/combiner les estimations spécifiques aux études

- Pour chaque identifiant de gène, identifier les études fournissant les informations d'abondance d'ARNm et de description de l'échantillon biologique pour la catégorisation des données.
- Pour chaque identifiant de gène, combiner l'estimation spécifique à l'étude et celles de toutes les études en utilisant une technique de méta-analyse. Enregistrer les résumés statistiques en résultant.
- Calculer la valeur p nominale du/des résumé(s) statistique(s) pour chaque identifiant de gène, et la corriger pour des tests multiples
- 7 Analyser, présenter et interpréter les résultats
 - Examiner la sensibilité des résultats pour les études individuelles en enlevant par exemple une étude et en faisant varier les sélections faites
 - Présenter le résultat statistique graphiquement pour les gènes d'intérêt
 - Analyses fonctionnelles des résultats (ex : des analyses d'enrichissement d'ontologie de gènes)
 - Si possible, valider en utilisant des technologies alternatives et/ou des échantillons différents
 - Considérer la force des preuves, les limitations et la possibilité de généralisation des conclusions

7.1 METHODES ET OUTILS POUR IDENTIFIER LES JEUX DE DONNEES ET EXTRAIRE LES DONNEES DES ETUDES

7.1.1 Identifier les jeux de données

La démarche classique est une recherche électronique de littérature afin de trouver des publications avec des données relatives à la question biologique, et ceci en utilisant les motsclés. Les sources et mots clés utilisés pour sélectionner les données à inclure dans la métaanalyse doivent être le plus exhaustif possible. Les principales sources consultées sont les bases de données bibliographiques informatisées telles que pubmed (https://www.ncbi.nlm.nih.gov/pubmed/), web of science (http://www.webofknowledge.com/) ou google scholar (https://scholar.google.fr/). D'autres bases de données généralistes existent : EMBASE, BIOSIS et PASCAL.

Pour le cas particulier des données de transcriptomiques, des bases spécialisées stockent les informations relatives aux échantillons, aux plans d'expérience de puces, et les données quantitatives d'abondance d'ADNc. Les 2 bases les plus documentées sont GEO (*Gene Expression Omnibus*, <u>https://www.ncbi.nlm.nih.gov/geo/</u>) du NCBI et ArrayExpress de EBI, (<u>https://www.ebi.ac.uk/arrayexpress/</u>). Tseng et *al.* (2012) ont aussi recensé des bases de données hébergées par des universités, instituts de recherche ou outils tels que Stanford Microarray Database (SMD), caArray at NCI, UPennRAD Database, UNC Microarray Database, Yale Microarray Database, MUSC Database et UPSCBASE ou enfin la base de données de Genevestigator (<u>https://genevestigator.com/gv/</u>).

Les bases de données de séquençage haut débit sont SRA (*Sequence Read Archive*, <u>https://www.ncbi.nlm.nih.gov/sra</u>) et ENA (*European Nucleotide Archive*, <u>https://www.ebi.ac.uk/ena</u>) (Figure 17)



Figure 17 : Lien entre les 4 bases de dépôt de données (source : <u>https://bioinfo-fr.net/bases-de-donnees-</u> <u>de-sequencage-geo-sra-ena-arrayexpress</u>)

Enfin, une autre possibilité est d'utiliser des bases de données préexistantes et spécialisées telles que Fat&MuscleDB (Tournayre et al., 2015) qui agrègent les données « omiques » liées à l'hypertrophie des tissus adipeux ou musculaires chez les mammifères. Tseng *et al.*, (2012) ont aussi listé des bases de données spécialisées relatives à une espèce, une pathologie ou le vieillissement, telles AGEMAP, Gene Age Nexus, Pancreatic Expression database, CYCLONET (relative au cycle cellulaire), ou les bases liées aux cancers Oncomine et Cancer Genome Workbench (CGWB).

Enfin, chaque étude sélectionnée doit être revue par un expert du domaine d'étude pour décider d'inclure ou d'exclure les données selon des critères à définir et à expliciter.

7.1.2 Les biais de publication

Il s'agit d'un biais lié au fait que les études positives sont plus souvent publiées alors que les études avec des résultats nuls ou négatifs ne sont pas toujours publiées. En outre, les études positives sont mieux publiées c'est-à-dire qu'elles sont dans des revues internationales et accessibles alors que les études négatives sont dans des revues peu accessibles, car peu référencées ou écrites dans une langue native.

Ce biais possible est critique pour la méta-analyse, car il influe et biaise la réponse générique attendue. Il existe 3 méthodes pour tester les biais de publication :

- La méthode de corrélation de rang (méthode de Begg)

- Une régression linéaire de l'effet estimé du traitement standardisé en fonction de la précision de l'estimation (méthode de Egger)
- Une régression de l'effet du traitement en fonction de la taille de l'échantillon (méthode visuelle utilisant les *funnel plot* ou graphiques en entonnoir). L'asymétrie visuelle du *funnel plot* cible le type de données manquantes, résultant de la non-publication de résultats négatifs (observé dans la Figure 18B). Ce type de graphique a aussi des limites, car l'asymétrie peut provenir d'un changement de réponse à un traitement ou de différences d'effet de traitement entre différents groupes ou études.

Il a cependant été montré par Macaskill, Walter, & Irwig (2001), qu'aucun test n'est vraiment efficace pour tester les biais dans la sélection des publications à inclure dans une méta-analyse, notamment du fait que les biais ont des origines multiples.



Figure 18 : Funnel plots montrant la relation entre la taille de l'effet et la taille des échantillons. A : en l'absence de biais de publication, le nuage de point attendu a une forme en entonnoir avec des points répartis symétriquement. B : un profil asymétrique du nuage de point suggère un biais de publications, et il n'y a que peu d'études négatives à faible effectif (figure provenant de (Dawson et al., 2016)).

Il existe d'autres type de biais de la méta-analyse, développé notamment dans le manuel de Cochrane, chapitre 8 (Higgins et Green, 2008).

7.1.3 Extraction des données

Les données issues des bases de données citées précédemment sont accessibles si elles ne sont pas protégées par une période d'embargo, et elles ont l'avantage de posséder des métadonnées (données qui décrivent le jeu de données : informations telles que l'espèce, la race, le nombre de répliquat biologiques utilisés...) uniformisées, ce qui facilite d'autant plus leurs réutilisations. Lorsque les données d'expression de gènes sont dans les publications ou dans les textes en supplément des publications, il faut d'abord regarder le nombre de données disponibles. Il y a habituellement peu de gènes décrits dans la publication, ce qui peut induire un biais en focalisant les gènes mis en avant par l'auteur et en excluant les autres gènes. Lorsque le nombre de données est suffisant, la récupération des données est réalisée à l'aide d'outils pour extraire les données de tableaux dans les PDF tels que tabula (https://tabula.technology/), PDFtables (https://pdftables.com/) or docparser (https://docparser.com/).

Les données disponibles en suppléments sont souvent plus fournies et parfois déjà au format de feuille de calcul, donc plus facilement manipulables. Cependant, ces données sont rarement complètes si elles sont issues de techniques telles que les puces à ADN ou le RNAseq (qui concernent souvent plusieurs milliers de gènes). Ce biais est à considérer lors de la sélection des données à inclure dans la méta-analyse.

7.2 **PREPARER, CATEGORISER ET METTRE A JOUR LES JEUX DE DONNEES**

Les publications portant sur la réutilisation de données (Rung et Brazma, 2013; Vaudel et al., 2016) mettent en avant que les données doivent être annotées au maximum. Les types d'annotation à considérer sont notamment le type de données : brutes versus traitées, le nom des auteurs, le plan expérimental, les publications liées à chaque jeu de données, des critères qualitatifs et quantitatifs liés aux échantillons.... Ces informations servent à coder et classer les études selon la question biologique à traiter. Pour le cas particulier des données omiques, les 2 étapes importantes dans la préparation des données sont la normalisation et la mise à jour des données. Il est à noter que les actions liées à la préparation des données sont extrêmement longues et le temps requis est très souvent sous-estimé lorsque l'on est novice dans le domaine.

7.2.1 Spécificité des données transcriptomiques en vue de leur méta-analyse

Les données transcriptomiques, qui étudient l'ensemble des ARNm (ARN messager) produit lors de la transcription, peuvent être produites avec différentes techniques telles que le séquençage d'ARN (en utilisant la technique du RNA-seq, provenant de *RNA-sequencing* en anglais) et les puces à ADN.

Les données transcriptomiques et notamment de puces à ADN et du RNAseq sont des abondances de milliers d'ARNm mesurées sur peu d'échantillons biologiques. Ces données

proviennent de méthodes de quantification par fluorescence, obtenues sur plusieurs puces pour une même étude. Ainsi, des étapes de normalisation de la fluorescence et des données obtenues sur plusieurs puces par étude sont nécessaires. Les données de puces à ADN issues de méthodes, puces ou plateformes différentes sont souvent normalisées avec des méthodes différentes.

Les données de transcriptomiques utilisées dans les publications ont plusieurs formes :

- Des données brutes ou des données archivées sont disponibles sur des bases de données telles que GEO, SRA, ArrayExpress, et ENA
- Des données archivées sont disponibles en données supplémentaires de l'article ou dans l'article lui-même.

La distinction entre données brutes et données archivées est importante, car les données brutes correspondent souvent aux données provenant directement de la plateforme avec peu de traitements de données préalables. Elles sont plus faciles de réutilisation, car le nouvel utilisateur choisit le traitement statistique et la normalisation à appliquer, les techniques étant en constante évolution et les buts de recherches pouvant varier d'un utilisateur à l'autre. Les données archivées sont des données qui ont déjà subi un traitement tel que la normalisation ou le calcul d'un ratio d'abondance. Il est alors plus difficile de réutiliser ces données dans une méta-analyse.

7.2.2 La normalisation

Les données ont besoin d'être normalisées selon une fonction de transformation afin de les rendre comparables. La normalisation est une étape indispensable à l'utilisation des données, quelle que soit la méthode d'obtention des données transcriptomiques (Nie, Wu, Culley, Scholten et Zhang, 2007). La normalisation est différente selon le type de puces (monochromatique ou pas), la plateforme d'analyse, ou l'année d'analyse.

La normalisation peut être réalisée via une feuille de calcul, avec des améliorations ammenées via des outils tels que EDGE (pour Extraction of Differential Gene Expression, un outil disponible librement sur demande (Leek, Monsen, Dabney et Storey, 2006) et également disponible en paquet R), mais peut également être réalisée via des paquets R dédiés tels que normalize (utilisé à la fois pour les données de puces à ADN et de RNAseq), Anapuce (pour les données de puces à ADN uniquement) ou intégré au paquet DEseq2 (utilisé dans le traitement des données de RNAseq).

Synthèse bibliographique - Méta-analyse

Lors de la préparation des données en vue de la méta-analyse, que nous traitons plus bas, différentes normalisations sont possibles. Ces normalisations s'effectuent avec des modèles basés sur des multi-array robustes (*model-based robust multi-array*, RMA), ou avec d'autres techniques de normalisation supprimant les contradictions entre les études et permettant la fusion directe des études, nommé XPN (pour *cross-platform normalization* (Shabalin, Tjelmeland, Fan, Perou et Nobel, 2008)), DWD (*Distance Weighted Discrimination* (Qiao, Zhang, Liu, Todd et Marron, 2010)) ou encore rGN (*ratio-adjust Gene-Wise* (Cheng, Shen, Song, Luo et Tseng, 2009)). Ces méthodes sont utilisables pour des puces provenant de même plateforme, mais les méthodes XPN et DWD sont également utilisées pour des puces à ADN provenant de plateformes différentes (Walsh, Hu, Batt et Santos, 2015).

7.2.3 Mise à jour des identifiants

Le problème majeur rencontré lors de la méta-analyse de données transcriptomiques est la diversité des identifiants utilisés pour un même gène. Cette diversité résulte en partie des différentes puces à ADN utilisées selon les laboratoires et les études. En effet, pour un même gène, les identifiants utilisés peuvent être ceux de GeneBank ou de Uniprot, d'autres sont les noms des gènes ou des identifiants créés. La diversité des identifiants est aussi due à la date de production de ces données et à la version du génome qui a permis l'annotation du gène selon la séquence ciblée. En effet, les jeux de données transcriptomiques « anciens » lient une séquence nucléique à un identifiant de gène selon la version connue du génome au moment de la genèse des puces à ADN utilisées. Lors de la mise à jour du génome, la séquence ciblée peut alors correspondre à un autre gène, ou correspondre à une région non codante. Enfin, des erreurs lors de l'annotation des sondes par rapport au génome causent aussi des inexactitudes d'identifiants (Gupta et al., 2016).

La mise à jour des jeux de données est souvent appelée réannotation. Cependant, la réannotation est en théorie un terme consacré aux génomes et non aux jeux de données transcriptomiques. Siezen et Van Hijum (2010) définissent la réannotation comme le processus permettant de mettre à jour un génome précédemment annoté (et non pas la mise à jour des jeux de données découlant de ce génome). Cependant, par abus de langage il n'est pas rare que le terme de réannotation soit utilisé pour parler de la mise à jour d'identifiants de gènes.

Il est déterminant de mettre à jour les identifiants de gènes en amont de la méta-analyse afin d'avoir des résultats fiables et de meilleure qualité, et pour maximiser les possibilités de trouver des gènes communs. En effet, dès 2005, Perez-Iratxeta et Andrade ont rapporté que pour une même puce murine (dont le 1^{er} séquençage du génome a été achevé en 2001, <u>https://www.broadinstitute.org/mouse/mouse-genome-project</u>), 5% des sondes étaient réassignées à de nouveaux gènes. Plus récemment, une mise à jour des puces ADN de poulet a revu 55% des identifiants. Ceci a permis d'améliorer la quantité et la qualité des annotations à la fois structurelles et fonctionnelles, et de modifier les résultats produits (Van den Berg, McCarthy, Lamont et Burgess, 2010). Enfin Arloth, Bader, Röh et Altmann (2015) ont montré que 25 % des gènes étaient mal annotés sur les puces à ADN humaines pourtant très souvent utilisées.

Il existe de nombreux outils de mise à jour de données transcriptomiques, soulignant l'enjeu et la volonté des scientifiques. Il serait difficile d'en faire une liste exhaustive d'autant que tous ces outils se basent sur l'alignement et la comparaison entre les séquences des sondes de la puce à ADN et celles (séquences géniques et intergéniques) du génome de l'espèce ciblée. L'alignement et la comparaison de séquences sont réalisés par un BLAST (Basic Local Alignement Search Tool) (Altschul, Gish, Miller, Myers et Lipman, 1990). Cependant pour réaliser ces mises à jour, il faut avoir accès aux séquences des ADN ciblant un ARNm/un gène.



Synthèse bibliographique – Méta-analyse

Figure 19 : Étapes d'obtention de la séquence nucléique d'un gène dans la base de données GEO sur la plateforme GPL4196 pour un utilisateur lambda

Ces informations sont généralement disponibles sur les bases de données telles que GEO, même si cela n'est en rien obligatoire lors du dépôt, mais fortement recommandé. Pour les données de puces à ADN, les séquences sont associées aux types de puces utilisées, qu'elles soient commerciales ou non. Cependant, les séquences ne sont pas toujours clairement affichées (ou disponibles) et plusieurs manipulations sont nécessaires pour les récupérer comme par exemple pour le GPL4196 (Figure 19). Un GPL (pour GEO Plateform) correspond aux diverses technologies, molécules et conventions d'annotation utilisées par les scientifiques. Le dépôt de celui-ci permet d'évaluer les données et de remonter aux identifiants et séquences ADN utilisés par la technologie (https://www.ncbi.nlm.nih.gov/geo/info/platform.html).

Afin de rendre accessible à des biologistes non programmeurs la mise à jour des identifiants de gènes à partir de séquences, une première liste non exhaustive d'outils a été établie par Siezen & Van Hijum (2010) et actualisé dans ce manuscrit (Tableau 9). Certains outils sont spécifiques à un organisme (ex. : SABIA pour les données d'annotation bactérienne) plus globaux tels quand d'autres sont que MAGPIE et ProbeLynx (http://koch.pathogenomics.ca/probelynx/). Ce dernier a été développé pour différentes espèces, dont le bovin (Roche et al., 2004). Cependant, cet outil n'est plus accessible depuis août 2013. De nombreux outils documentés tels que ArrayIDer, (B. H. van den Berg, Konieczka, McCarthy et Burgess, 2009), ReMOAT (Barbosa-Morais et al., 2009) ou ProbeLynx (Roche et al., 2004) ne sont plus maintenus.

| Pipeline | Organization | Description | Access/distribution | Reference |
|---------------------|--|--|---|---|
| IGS | University of Maryland | A FREE resource for genomics researchers and educators bringing advanced bioinformatics tools to the lab bench and the classroom | http://www.igs.umarylan d.edu/research/bioinform atics/analysis/index.php Free service | None |
| MaGe/Mic roScope | GENOSCOPE | Magnifying Genomes: microbialgenome annotation system | http://www.genoscope.cn s.fr/agc/mage Free service | Vallenet et al. (2006)et al. (2009) |
| RAST | Fellowship for Integration of Genomes (FIG) | The RAST Server: Rapid Annotations using Subsystems Technology based on the Seed | http://rast.nmpdr.org/ Free to use service | Aziz et al. (2008) |
| AGMIAL | INRA, France | An annotation strategy for prokaryote genomes as a distributed system | http://genome.jouy.inra.f r/agmial/ | Bryson et al. (2006) |

Tableau 9 : Liste non exhaustive d'outils pour la mise à jour des identifiants de gènes à partir de séquences (tableau adapté de Siezen & Van Hijum (2010)).

| | | | Open source license | |
|--------------------------------------|---|--|---|--|
| DIYA | DIY Genomics Consortium | A bacterial annotation pipeline for any genomics lab | https://sourceforge.net/pr ojects/diyg/ Free to use, stand-alone software | Stewart et al. (2009) |
| SABIA | LNCC, Brazil | Bacterial annotation system | http://www.sabia.lncc.br/ Free to use, stand-alone software | Almeida et al. (2004) |
| Restauro-G | Institute for Advanced Biosciences, Keio University | À Rapid Genome Re-Annotation System for Comparative Genomics | http://restauro- g.iab.keio.ac.jp/ Software distributed under the GNU public license | Tamaki et al. (2007) |
| ATUCG system | Universidade Federal do Rio Grande do Sul, Brasil | Agent-based environment for automatic annotation of Genomes | None Software should be requested at authors | Nascimento and Bazzan (2005) |
| Taverna: annotation of genomes | University of Manchester | Interactive genome annotation pipeline. | http://www.taverna.org.u k/introduction/taverna- in- use/annotation/annotatio n-of-genomes/ | Hull et al. (2006) |
| KAAS | Kyoto Encyclopedia for Genes and Genomes (KEGG) | KEGG automated annotation service for metabolic pathways | http://www.genome.jp/to ols/kaas/ Free to use service | Moriya et al. (2007) |
| Re- annotator | | Re-annotation pipeline for gene expression microarrays that will bring probe annotations up-to- date (for mouse and human) | https://sourceforge.net/pr ojects/reannotator/ Free to use service | (Arloth et al., 2015) |
| SigReannot -mart | EADGENE, INRA de Toulouse | SigReannot-mart is a query environment populated with regularly updated annotations for different oligo sets. It has be mainly used on farm and aquaculture species | http://sigreannot- mart.toulouse.inra.fr Free to use service | (Moreews, Rauffet, Dehais et Klopp, 2011) |

Parmi les outils maintenus, **Re-annotator** est disponible ligne en (https://sourceforge.net/projects/reannotator/) ou utilisable en version locale sur un ordinateur. Il a été comparé notamment à un outil plus ancien, ReMOAT (Re-annotation and Mapping for Oligonucleo- tide Array Technologies) qui n'est plus maintenu et qui était un outil de référence pour la mise à jour des puces Illumina (Barbosa-Morais et al., 2009). ReMOAT se base sur le génome de référence uniquement tandis que Re-annotator se base sur les séquences RefSeq et sur le génome entier, en considérant de possibles décalages. Une différence de 16% est observée entre les annotations de Re-annotator et ReMOAT, avec un nombre de sondes susceptibles d'être réannotées plus important dans le cas de Re-annotator (Arloth et al., 2015). L'outil Reannotator, permet *à priori* la mise à jour de séquences/identifiants de 2 espèces (l'humain et la souris), et demande des connaissances dans plusieurs langages de programmation, tels que le Shell et le Perl. Ceci peut constituer une limite d'utilisation pour les scientifiques non formés à l'informatique/bio-informatique.

Un autre outil est toujours en activité : SigReannot-mart (<u>http://sigreannot-mart.toulouse.inra.fr</u>). Cet outil est utilisable facilement par un scientifique, car il demande peu de connaissance en bio-informatique. Il est spécialisé pour la mise à jour des identifiants de gènes de diverses puces à ADN à partir de séquences de plusieurs espèces d'animaux de rente: le cheval, la poule, le lapin ou encore le bovin (Moreews et al., 2011).

Enfin, parmi les derniers outils préparatoires à la mise à jour des séquences stockées au format GenBank, le paquet R AnnotationBurstR (Borstein et O'Meara, 2018) a été mis en ligne mi-2018. Il réalise le téléchargement des séquences, au format FASTA, de gènes contenus sur GenBank.

7.3 CHOISIR UNE METHODE DE META-ANALYSE POUR IDENTIFIER DES DIFFERENTIELS D'ABONDANCE DE GENE

Tseng *et al.* (2012) a catégorisé 328 des 333 publications de méta-analyses de résultats. Comme l'indique la Figure 20, 66 % des études ont pour objectif d'identifier des gènes différentiellement exprimés. Pour atteindre cet objectif, quatre méthodes sont majoritairement utilisées comme l'indique la revue publiée par Tseng *et al.* (2012). Ces méthodes sont la fusion directe des données brutes, la combinaison de rang, l'effet de taille, ou la valeur de probabilité (nommée valeur p) et sont présentées dans les Figure 14 et 21.



Figure 20 : Objectifs des méta-analyses sur des données de puces à ADN déclarés dans 328 publications (adapté de Tseng et al. 2012). Quatre types d'objectifs, avec des proportions (en %) variables, ont été catégorisés, le majeur étant la recherche de gènes différentiellement exprimés.



Figure 21 : Méthode de méta-analyses utilisées dans les 191 méta-analyses cherchant à mettre en évidence les gènes différentiellement exprimés (adaptée de Tseng et al. 2012).

62

7.3.1 Fusion directe

Cette méthode, également appelée « méga-analyse », consiste à normaliser les données pour pouvoir les fusionner en un seul et même jeu de données.

Après une normalisation (les différentes méthodes utilisées sont décrites dans le chapitre 7.2.1), les données sont assemblées directement. Des procédures statistiques standards pourront être appliquées pour détecter les gènes DE, pour le regroupement (*clustering*) ou la classification.

Cependant, lors de cette fusion d'étude, une attention est à apporter à la normalisation qui n'enlève pas toutes les contradictions entre études. D. R. Goldstein, Delorenzi, Luthi-Carter, & Sengstag, (2011) ont montré que la méthode RMA de normalisation n'a pas permis d'éliminer les effets de lot (*batch effects*), même avec 2 études provenant du même laboratoire et de la même plateforme Affymetrix, mais produites à deux moments différents.

7.3.2 Combinaison de rang

La combinaison de rang consiste à analyser les résultats d'abondance des gènes et à leur donner un rang selon leur abondance et la valeur de la probabilité montrant que la différence entre 2 conditions n'est pas due au hasard.

Pour chaque gène, la somme ou le produit de l'abondance des ARNm est calculé. Puis, pour les données génomiques, et plus particulièrement pour les données de puce à ADN, un test de Student est effectué pour identifier chaque gène avec une expression différentielle entre deux conditions. Les gènes ont ensuite un rang attribué selon la valeur de probabilité ou valeur p (ou autres valeurs calculées telles que la q-value). Sur n gènes testés, le rang le plus grand (n) est attribué au gène ayant la plus petite valeur p et la plus forte différence d'abondance entre les traitements testés (Kolde, Laur, Adler et Vilo, 2012).

Le rang est calculé pour chaque gène dans chaque étude. Avant de combiner les études, les rangs bruts doivent être ajustés par le nombre maximum de variables testées (n_{max}) dans l'ensemble des études combinées. Ainsi, pour l'étude i ayant un nombre de variables testées n_i , les rangs sont multipliés par n_{max}/n_i (Zintzaras et Ioannidis, 2008).

Les rangs calculés sont ensuite comparés par comparaison de liste des gènes les plus différentiellement exprimés (DeConde et al., (2006). Cette comparaison des différentes études permet de tirer des conclusions à partir des gènes communs entre les différentes études pour un seuil spécifique. Un exemple est donné dans la Figure 22, avec un seuil établi aux 25 premiers gènes.

| Rank | Luo | Welsh | Dhana | True | Singh |
|------|---------|---------|---------|---------|---------|
| 1 | HPN | HPN | OGT | AMACR | HPN |
| 2 | AMACR | AMACR | AMACR | HPN | SLC25A6 |
| 3 | CYP1B1 | OACT2 | FASN | NME2 | EEF2 |
| 4 | ATF5 | GDF15 | HPN | CBX3 | SAT |
| 5 | BRCA1 | FASN | UAP1 | GDF15 | NME2 |
| 6 | LGALS3 | ANK3 | GUCY1A3 | MTHFD2 | LDHA |
| 7 | MYC | KRT18 | OACT2 | MRPL3 | CANX |
| 8 | PCDHGC3 | UAP1 | SLC19A1 | SLC25A6 | NACA |
| 9 | WT1 | GRP58 | KRT18 | NME1 | FASN |
| 10 | TFF3 | PPIB | EEF2 | COX6C | SND1 |
| 11 | MARCKS | KRT7 | STRA13 | JTV1 | KRT18 |
| 12 | OS-9 | NME1 | ALCAM | CCNG2 | RPL15 |
| 13 | CCND2 | STRA13 | GDF15 | AP3S1 | TNFSF10 |
| 14 | NME1 | DAPK1 | NME1 | EEF2 | SERP1 |
| 15 | DYRK1A | TMEM4 | CALR | RAN | GRP58 |
| 16 | TRAP1 | CANX | SND1 | PRKACA | ALCAM |
| 17 | FMO5 | TRA1 | STAT6 | RAD23B | GDF15 |
| 18 | ZHX2 | PRSS8 | TCEB3 | PSAP | TMEM4 |
| 19 | RPL36AL | ENTPD6 | EIF4A1 | CCT2 | CCT2 |
| 20 | ITPR3 | PPP1CA | LMAN1 | G3BP | SLC39A6 |
| 21 | GCSH | ACADSB | MAOA | EPRS | RPL5 |
| 22 | DDB2 | PTPLB | ATP6V0B | CKAP1 | RPS13 |
| 23 | TFCP2 | TMEM23 | PPIB | LIG3 | MTHFD2 |
| 24 | TRAM1 | MRPL3 | FMO5 | SNX4 | G3BP2 |
| 25 | YTHDF3 | SLC19A1 | SLC7A5 | NSMAF | UAP1 |

Figure 22 : les 25 gènes les plus surexprimés. Le rang est celui observé pour chacune des 5 études analysées par DeConde et al., (2006).

7.3.3 La taille de l'effet

La taille de l'effet (*effect size*) est un moyen de quantifier la différence entre deux groupes, ce qui présente de nombreux avantages par rapport à l'utilisation de tests de significativité statistique uniquement. La taille de l'effet met l'accent sur la taille de la différence

ce qui évite de la confondre avec la taille de l'échantillon (Borenstein et al., 2009). La taille de l'éffet est calculée comme suit :

[Taille de l'effet]= [Moyenne du groupe traité]-[Moyenne du groupe contrôle] [Écart type du groupe contrôle]

L'écart type est une mesure de la dispersion d'un ensemble de valeurs. Ici, il s'agit de l'écart type de la population à partir de laquelle deux groupes ont été étudiés. Toutefois, dans la pratique, cela n'est presque jamais connu et doit donc être estimé à partir de l'écart type du groupe contrôle ou à partir de valeurs communes entre les deux groupes (Sullivan et Feinn, 2012).

D'autres types de taille d'effet peuvent être calculés : le rapport de chance (*odds ratio*, calculé en divisant les rapports de cote des risques absolus du groupe 1 par celle du groupe 2). Pour les données binaires, un calcul de ratio de risque (calculé en divisant la probabilité de résultats du groupe 1 par celle du groupe 2) compare la probabilité que les résultats soient reproduits d'une étude à l'autre.

7.3.4 Combinaison de taille de l'effet

L'un des modèles de la combinaison de la taille de l'effet (*Combining effect sizes*) est basé sur la supposition que la taille de l'effet standardisé est comparable entre études. Dans cette catégorie, les modèles à effets fixes et à effets aléatoires (respectivement FEM et REM) sont les approches les plus populaires.

7.3.4.1 Effet fixe de la combinaison de taille de l'effet (FEM)

Pour le modèle à effet fixe, il est supposé que la vraie taille de l'effet de chaque étude est identique et que la variation de taille de l'effet observée entre les études est due à une estimation d'erreur entre les études. Ainsi, toutes les différences observées sont considérées comme provenant d'une erreur expérimentale ou d'une erreur d'échantillonnage de la population (Borenstein, Hedges, Higgins et Rothstein, 2010).

Ce modèle doit être choisi si 2 conditions sont rencontrées :

 Les études ont été construites avec des modalités expérimentales identiques (même type d'échantillons, mêmes sondes, même chercheur, même laboratoire...) Le but est de calculer la taille de l'effet commun aux études et ce résultat ne pourra pas être généralisé au-delà de la population choisie pour la méta-analyse.

Dans le cas d'un effet fixe, l'assignation à un groupe d'une variable ou à un autre (que ce soit A, B, C, D ou E dans l'exemple présenté dans la Figure 23) n'a pas d'impact sur le score final et donc chacune des études partage une même taille d'effet. C'est la taille relative du groupe (ici N ou relative weight) qui est prise en compte pour calculer la taille de l'effet (le losange en bas de la Figure 23). Donc l'étude C, dont l'effectif est supérieur à celui des autres études influence fortement le calcul de taille de l'effet.



Figure 23 : Illustration de la manière de calculer l'effet fixe de la taille de l'effet par Borenstein et al. (2010). La taille de l'effet de chaque étude est représentée par un carré et par l'intervalle de confiance.

7.3.4.2 Effets aléatoires de la combinaison de taille de l'effet (REM)

Pour les modèles avec un effet aléatoire, on suppose que chaque étude contient un effet aléatoire intra-étude conduisant à l'incorporation d'un effet inconnu dans la méta-analyse. Cet effet aléatoire est dû à l'hétérogénéité provenant du croisement d'études dans le modèle. Le but n'est donc pas d'estimer une vraie taille d'effet, mais d'estimer la moyenne de distribution des tailles d'effet. Comme chaque étude donne des informations sur une taille d'effet différente, l'idée est d'estimer la moyenne de la distribution de l'effet de taille. Ainsi, toutes les études auront le même poids et la même variance, car chacune d'elles apporte une information sur un effet. Ceci le distingue du modèle à effet fixe dans lequel une étude plus grande que les autres aurait un poids et une variance plus importante (Borenstein et al., 2010).

Dans notre exemple (Figure 23), avec les mêmes valeurs que celles utilisées pour l'exemple de calcul de l'effet fixe, l'assignation d'une valeur à un groupe est spécifique. Cependant, on observe que l'écart du poids des études entre les études à faible et fort effectifs est plus faible pour les modèles à effet aléatoire que pour les modèles à effet fixe (les études avec un petit effectif ont plus de poids dans le calcul final comparé à l'effet fixe).



Figure 24 : Illustration de la manière de calculer l'effet aléatoire de la taille de l'effet par Borenstein et al. (2010). La taille de l'effet de chaque étude est représenté par un carré et par l'intervalle de confiance.

Dans une méta-analyse utilisant l'effet aléatoire, l'hypothèse est que les études ont assez de points en communs pour la synthèse de l'information, et qu'il n'y a pas de raison de supposer qu'elles sont identiques dans le sens où elles auraient toute la même taille de l'effet. Les résultats de la méta-analyse pourront donc être généralisés à une population plus importante (Borenstein et al., 2010).

7.3.5 Les valeurs p

La valeur p (valeur de probabilité ou p-value pour *probability value*) correspond à la probabilité que la différence observée entre 2 groupes soit due à la chance (l'hypothèse nulle ou H0, et qu'il n'y a en fait pas de différence vraie entre les 2 groupes). Un résultat considéré comme statistiquement significatif est un résultat qui serait faux si l'hypothèse nulle était vérifiée. Une valeur seuil α (généralement fixée à 5%) correspond à la valeur sous laquelle le résultat est jugé statistiquement significatif et pour laquelle l'hypothèse nulle est rejetée. Cependant, si la valeur p est au-dessus de 5%, l'hypothèse nulle ne peut pas être rejetée (Aschwanden, 2015).

7.3.6 Combinaison de valeur p

Pour les méta-analyses de puces à ADN, la méthode majoritairement utilisée selon Tseng et al. (2012) est la combinaison des valeurs p qui présente plusieurs avantages majeurs :

- Une simplicité et une extensibilité à différents types de variables
- Une standardisation sur une échelle commune des associations provenant d'études génomiques
- Une quantité minimale d'informations provenant des études

Avec des résultats de valeurs p, il est possible de faire des diagrammes de Venn ou d'appliquer des méthodes de comptage de votes. Tandis que les diagrammes de Venn sont des outils de visualisation utiles de la distribution des DEGs dans les études individuelles, la méthode de comptage de vote consiste à compter le nombre d'études avec des valeurs p sous un seuil convenu (P<0.05). Cependant, étant donné la perte d'information induite par le comptage de vote, cette dernière méthode est à éviter.

Il existe plusieurs méthodes de combinaison de valeurs-p, avec différentes méthodes telles que les méthodes de Fisher, de Pearson, de George, d'Edgington, de Stouffer, de Tippett (Heard et Rubin-Delanchy, 2018), de Wilkinson, de Lipták, de Lancaster ou encore de Mudholkar et George (Loughin, 2004). Elles peuvent également être nommées selon la manière dont sont combinées les valeurs p (Tableau 10):

- minimum
- chi carré (χ^2) (utilisé dans la méthode de Fisher et de Lancaster)
- maximum (utilisé dans la méthode de Wilkinson)
- uniforme (utilisé dans la méthode d'Edgington)
- logistique (utilisé pour la méthode de Mudholkar et George)
- normale (utilisé dans la méthode de Lipták et Stouffer)

Nous ne décrirons pas chacun de ces modèles dont les descriptions ont été rapportées par Loughin, (2004), Bhandary & Zhang, (2011) et Heard & Rubin-Delanchy, (2018). Une synthèse des modèles et méthodes de combinaison de valeurs p est donnée dans le Tableau 10.

| Modèle | Année | Type de combinaison des valeurs p | Référence |
|----------------------|-------|---|---|
| Pearson | 1933 | Chi-square | (Pearson, 1933) |
| Stouffer | 1949 | Standard normal | (Stouffer, Suchman, Devinney, Star, & Williams, 1949) |
| Tippett | 1931 | Minimum | (TIPPETT, 1931) |
| Liptak | 1958 | Standard normal | (Liptak, 1958) |
| Fisher | 1934 | Chi-square | (Fisher, 1925) |
| Edgington | 1972 | Uniforme | (Heard et Rubin-Delanchy, 2018) |
| Lancaster | 1961 | Chi-square | (Lancaster, 1961) |
| Mudholkar and George | 1979 | Logistique | (Mudholkar et George, 1979) |

Tableau 10 : Principaux modèles de combinaison de valeurs p

7.3.7 Taille de l'effet versus valeurs p

Les valeurs p informent sur l'existence d'un effet tandis que la taille de l'effet quantifie la différence entre deux groupes et est considérée comme une mesure exacte de l'importance de la différence (Sullivan et Feinn, 2012). Ces deux approches partent donc des mêmes données, mais ne cherchent pas à répondre à la même question (Borenstein et al., 2009, p. 299).

L'utilisation des valeurs p plutôt que des tailles d'effet, bien que majoritaire pour les méta-analyses, a été critiquée notamment par (Sullivan et Feinn, 2012) et (Borenstein et al., 2009, p. 297) avec les arguments suivants :

- La taille de l'effet donne plus d'informations que la méthode des valeurs p puisqu'elle indique la significativité et l'ordre de grandeur de l'effet. Cependant, il est nécessaire d'avoir accès aux données brutes afin de calculer la taille de l'effet (Zollinger, Davison et Goldstein, 2015).
- Les résultats des analyses avec les valeurs p sont souvent mal interprétés (Higgins & Green, 2008, p. 372) : d'une part, les valeurs p très petites sont interprétées comme un effet très important du traitement. Ce n'est évidemment pas le cas, la valeur p répondant uniquement à la question « oui il y a une différence », mais ne répond pas à la question « cette différence est de x fois). D'autre part, la 2^e mauvaise interprétation de la valeur p est qu'une valeur p plus grande que le α (habituellement α=0.05) est considérée

comme une preuve que le traitement/l'intervention n'a pas d'effet alors que l'interprétation correcte est qu'il n'y a pas de preuve que le traitement ait un effet. Afin d'éviter ces biais, le manuel de Cochrane pour les revues systématiques (Higgins et Green, 2008) recommande les analyses de l'intervalle de confiance et de l'estimation de l'effet.

Pour conclure, différentes stratégies statistiques s'appliquent à la méta-analyse de données transcriptomiques, la plus courante toutes espèces confondues, étant la combinaison de valeur p. Les rares méta-analyses réalisées à partir de données « omiques » bovines ont utilisé la combinaison de valeur p pour identifier les signatures transcriptomiques de la lactation (Farhadian, Rafat, Hasanpur et Ebrahimie, 2018), ou de la maturité du follicule ovarien et du développement des ovocytes (Khan et al., 2016). Baron et al., (2011) ont rapporté les gènes impliqués dans le développement du muscle et co-exprimés similairement entre les différentes espèces par une méthode de comparaison de cluster, qui s'apparente à de la comparaison de rang décrite plus haut.

7.4 ANALYSER, PRESENTER ET INTERPRETER LES RESULTATS

Les résultats attendus après des méta-analyses de données transcriptomiques sont des listes de gènes différentiellement exprimés selon les conditions comparées. Cette liste doit ensuite être interprétée pour pouvoir répondre à la question biologique posée lors de la réalisation de la méta-analyse. Ces analyses utilisent souvent des identifiants de gènes et ont pour objectifs de renseigner sur les fonctions biologiques impliquant les gènes mis en avant par la méta-analyse, le rôle central des gènes dans des voies de régulation, les gènes qui sont corégulés et enfin, dans le contexte de la recherche de biomarqueurs peu invasifs, les gènes qui codent pour des protéines sécrétées dans les fluides biologiques. Les analyses dites fonctionnelles qui répondent à ces objectifs sont des enrichissements d'annotation selon l'ontologie des gènes, la construction de réseaux et l'analyse de leur topologie, la prédiction de la sécrétion ou de la localisation des protéines codées par les gènes.

7.4.1 Ontologie de gènes et enrichissement

L'ontologie de gènes (*Gene ontology, GO*, en anglais) est un dictionnaire hiérarchisé permettant de décrire les propriétés des produits géniques selon trois axes : les processus biologiques (BP pour *biological process*), la localisation cellulaire (CC pour *cellular components*) et les fonctions moléculaires (MF pour *molecular function*). Les gènes/protéines sont « annotés » par un ou plusieurs termes GO, car ils jouent un rôle dans une ou plusieurs fonctions biologiques. L'analyse GO renseigne donc sur les gènes impliqués ou faisant partie d'une même catégorie (Huntley, Sawford, Martin et O'Donovan, 2014) et propose de décrire les relations entre les GO. En outre le calcul de l'enrichissement GO identifie les termes GO et les gènes/protéines associées, statistiquement surreprésentées dans un jeu de données (Buza, Mccarthy, Wang, Bridges et Burgess, 2008). Les termes de GO (ou la voie métabolique) avec la valeur p la plus faible sont les termes qui sont considérés comme les plus enrichis.

Les « outils » majeurs pour réaliser les analyses d'ontologie de gènes sont : AmiGO (Carbon et al., 2009), QuickGO (Huntley et al., 2009), UniProt-GOA (Huntley et al., 2015), DAVID (Huang, Sherman et Lempicki, 2009), GoMiner (Feng et al., 2003), FatiGO (Al-Shahrour, Díaz-Uriarte et Dopazo, 2004), GOrilla (Eden, Navon, Steinfeld, Lipson et Yakhini, 2009), PANTHER (J. E. Chen et Glover, 2016), WebGestalt (B. Zhang, Kirov et Snoddy, 2005) ou ProteINSIDE (Kaspric, Picard, Reichstadt, Tournayre et Bonnet, 2015) développé pour analyser des gènes/protéines de ruminants. Ces outils utilisent souvent des identifiants de gènes ou de protéines pour une ou plusieurs espèces. Par exemple, ProteINSIDE, comparativement à David et AGbase, annote entre 14 et 99 % plus d'identifiants bovins, ovins ou caprins. La plupart des outils calculent l'enrichissement des GO, d'autres calculent en plus l'enrichissement des voies métaboliques telles que celles classifiées par KEGG.

7.4.2 Co-expression des gènes

L'objectif est d'analyser les relations entre les gènes, puis d'observer l'impact d'un traitement sur ces relations. Ces relations peuvent être des corrélations deux à deux de l'abondance d'ARNm, qui servent alors de base pour regrouper ou classer les gènes selon les résultats de corrélations ou de co-expression. Les méthodes statistiques les plus utilisées pour calculer les relations sont des méthodes de classification hiérarchique (telle que la classification hiérarchique ascendante) ou de partitionnement de données (data clustering en anglais) non

hiérarchisé (par l'utilisation d'Analyse en Composantes Principales, ACP ou par l'utilisation des k-means) (Karimpour-Fard, Epperson et Hunter, 2015). Les méthodes de classification ont pour objectifs de fusionner successivement les données en clusters (groupes) puis les clusters entre eux selon leur degré de similarité (distance). La méthodes de classification cherche à faire des classes telles que 1) les différences intra-classes soient minimales pour obtenir des clusters ; 2) les différences inter-classes soient maximales afin d'obtenir des sous-ensembles bien différenciés.

Les outils classiques de traitement de données Minitab tel que (https://www.minitab.com/fr-fr/), **PQStat** (https://pqstat.pl/), XLSTAT ou encore (https://www.xlstat.com/fr/) offrent la possibilité de faire des analyses de classification ou de clustering. Parmi les outils utilisés en analyses omiques, on retrouve notamment R (avec les paquets ade4, heatmap2 ou FactoMineR par exemple) et Permutmatrix (Caraux et Pinloche, 2005) développé pour les biologistes non programmeurs.

7.4.3 Réseaux moléculaires

Un réseau d'interactions de molécules peut être constitué avec des protéines, des gènes ou de petites molécules telles que les lipides ou les drogues. Ces gènes ou protéines constituent les nœuds (*node* en anglais), reliés entre eux (ou non) deux à deux par une relation donnée, nommée arrête (*edge* en anglais). Les liaisons entre ces nœuds sont par exemple des interactions physiques identifiées expérimentalement, des liens fonctionnels supposés ou vérifiés ou des corrélations de niveaux (Li, Yang, Zhang et Jiao, 2019; J. Zhang, Zhong, Lin et Wang, 2019). Lorsque la relation n'est pas orientée, le réseau est de type non orienté. Lorsque la relation a un sens biologique, il s'agit de réseaux orientés, ils nécessitent une plus grande connaissance de la relation entre 2 nœuds que des réseaux non orientés.

Parmi les bases de données qui répertorient les liaisons entre molécules, certaines sont spécifiques à une espèce (DroID, base de données d'interaction spécifique à la drosophile) tandis que d'autres sont spécialisées dans un type d'interaction (MatrixDB spécialisée dans les interactions extracellulaires). Il existe des bases de données non spécialisées dont les plus courantes sont décrites dans le Tableau 11. La plus connue est STRING qui permet notamment

de tracer un réseau de protéines ou de gènes en se basant sur leurs interactions expérimentales ou issues d'analyses de texte.

Tableau 11 : Principales bases de données (non spécifique à un organisme) des interactions entre molécules.

| Base de | Description | Référence |
|-----------|---|--|
| PSICQUIC | Proteomics Standards Initiative Common QUery InterfaCe, est un effort communautaire pour standardiser la manière d'accéder et de récupérer les données provenant des bases de données d'interactions moléculaires <u>http://www.ebi.ac.uk/Tools/webservices/psicquic/view/main.x</u> <u>html</u> | (Aranda et al., 2011) |
| IMEX | Contient les données de nombreuses bases de données publiques d'interaction telles que IntAct, BioGRID, Uniprot (https://www.imexconsortium.org/) | (Orchard et al., 2012) |
| IntAct | Base de données d'interactions moléculaires qui est en open source et qui fournit également des outils d'analyse. Toutes les interactions sont issues de la littérature ou de la soumission directe par l'utilisateur (<u>https://www.ebi.ac.uk/intact</u> /) | (Orchard et al., 2014) |
| STRING | Collecte et intégration des données en consolidant les données prédites d'association protéine-protéine pour un grand nombre d'organismes. Les associations incluent les interactions directes (physiques) aussi bien que les interactions indirectes (fonctionnelles), si celles-ci sont spécifiques et significatives biologiquement. (https://string-db.org/) | (Szklarczyk et al., 2017; von Mering et al., 2003) |
| Pathguide | Un abrégé des données pour les interactions selon les espèces. Contient des informations à propos de 702 ressources liées à des voies biologiques et aux ressources liées à des données d'interactions moléculaires. <u>http://www.pathguide.org/</u> | (Bader, 2006) |

La topologie du réseau de molécule fournit des informations très importantes telles que la densité du graphe qui est le quotient entre le nombre d'arrêtes et le nombre de paires de nœuds. Elle informe aussi des nœuds centraux, essentiels à la stabilité du réseau. Pour identifier ces nœuds centraux, on calcule la fréquence des plus courts chemins liés à une molécule (*closeness centrality*) et/ou la fréquence d'apparition du nœud entre des groupes de molécules (*betweenness centrality*). Cela permet d'identifier les molécules régulatrices les plus importantes du jeu de données testé. Des outils permettent de visualiser ces réseaux tels que STRING (Szklarczyk et al., 2017; von Mering et al., 2003) et ProteINSIDE (Kaspric et al., 2015).

Il est également possible de représenter directement les voies métaboliques et les cascades de régulation mises en évidence dans le jeu de données. Ceci peut se faire via des outils de visualisation de données biologiques tels que Cytoscape (interface gratuite (Shannon, 2003)) et Pathway Studio (outils sous licence, Yuryev, Kotelnikova et Daraselia, 2009).

7.4.4 Sécrétion des protéines

Dans un contexte de recherche de biomarqueurs peu invasifs, la recherche de molécules dans des fluides biologiques est un objectif en médecine humaine, mais aussi pour la mesure de caractères phénotypiques chez les animaux, dont les animaux de rente. Les méthodes « omiques » appliquées aux fluides identifient directement et expérimentalement ces biomarqueurs. De tels biomarqueurs peuvent être prédits au niveau protéomique ou transcriptomique par exemple. En effet, environ 10% du génome humain conduit à la synthèse de protéines sécrétées (Skalnikova, Motlik, Gadher et Kovarova, 2011). L'ensemble des protéines sécrétées d'une cellule est appelé sécrétome. Des méthodes de prédiction de ce sécrétome ont été développées (Klee et Ellis, 2005), nécessitant les séquences protéiques pour prédire les peptides signaux. Il existe au moins 5 voies de sécrétion des protéines chez les eucaryotes. Les protéines néo-synthétisées avec un peptide signal N-terminal sont orientées vers la voie classique de sécrétion incluant leur translocation dans la lumière du réticulum endoplasmique puis leur transport dans l'appareil de Golgi et leur sécrétion par exocytose. Les quatre autres voies de sécrétion « non classiques » impliquent les transporteurs membranaires, le recyclage de l'endosome, le bourgeonnement de la membrane conduisant à la formation de vésicules ou des exosomes contenant notamment des protéines ou les translocations de protéines des feuillets externes de la membrane plasmatique (Kaspric et al., 2015). Les outils les plus utilisés pour la mise en évidence de protéines sécrétées par l'une de ces 5 voies sont décrits dans le Tableau 12.

| Outils | Type de prédiction | Référence | Site |
|-----------------|--|--|---|
| Prédiction de p | peptides signaux de la voie | classique | - |
| SignalP | Prédiction de signal peptide | (Petersen, Brunak, Von Heijne et Nielsen, 2011) | http://www.cbs.dtu.dk/s ervices/SignalP/ |
| PrediSi | Prédiction de signal peptide | (Hiller, Grote, Scheer, Münch et Jahn, 2004) | http://www.predisi.de/ |
| SigCleave | Prédiction de signal peptide | (Rice, Longden et Bleasby, 2000) | http://emboss.bioinform atics.nl/cgibin/ emboss/sigcleave |
| Phobius | Prédiction de signal peptide | (Käll, Krogh et Sonnhammer, 2007) | http://phobius.sbc.su.se/ |
| ProteINSIDE | Prédiction de signal peptide et d'ontologie de gènes associées | (Kaspric et al., 2015) | http://www.proteinside.org/ |
| Autres voies de | e sécrétion | | |
| TargetP | Prédiction de localisation cellulaire | (Emanuelsson, Nielsen, Brunak et Von Heijne, 2000) | http://www.cbs.dtu.dk/s ervices/TargetP/ |
| PSORTb | Prédiction de localisation cellulaire | (Yu et al., 2010) | http://www.psort.org/ps ortb/ |
| SecretomeP | Prédiction de protéine sécrétée par une autre voie | (Bendtsen, Jensen, Blom, Von Heijne et Brunak, 2004) | http://www.cbs.dtu.dk/s ervices/SecretomeP/ |
| ProteINSIDE | Prédiction de localisation cellulaire et d'ontologie de gènes associés | (Kaspric et al., 2015) | http://www.proteinside.org/ |

Tableau 12 : Exemple d'outils de prédiction des protéines sécrétées, adapté de (Caccia, Dugo, Callari et Bongarzone, 2013)

La connaissance de ces protéines sécrétées (ainsi que leurs actions) permet de cibler plus spécifiquement certains gènes/protéines qui seraient susceptibles d'avoir une action non seulement au niveau de la cellule, mais également à l'échelle du tissu ou d'un organisme.

Il est à noter cependant qu'aucune base spécifique au sécrétome de ruminant n'existe. Cependant, la plupart des outils se basent sur la séquence pour rechercher les motifs de peptides signaux. La plupart des protéines indiquées comme sécrétées chez le bovin, sur la base de données Uniprot par exemple, l'ont été manuellement par des experts ou par similarité de séquence entre espèces.

8 OUTILS DE META-ANALYSE POUR LES DONNEES DE TRANSCRIPTOMIQUE

L'essentiel des outils disponibles sont des scripts pour le logiciel R. Le logiciel R, accessible gratuitement, et à code ouvert (*open source*, répondant aux critères d'accès au code source, de libre redistribution et de création de travaux dérivés), est le fruit de la collaboration entre des programmeurs. Il recense pour le moment plus de 14849 paquets dans le référentiel de paquet CRAN (*Comprehensive R Archive Network*, <u>https://cran.r-project.org/</u>, le 27/08/19 à 9h) (R Core Team, 2018). Il est également possible de trouver des packages disponibles sur Bioconductor (<u>https://www.bioconductor.org/</u>).

En 2017, Polanin, Hennessy, & Tanner-Smith ont dénombré pas moins de 63 paquets spécifiques à la méta-analyse. En 2019 (le 27/08), il y avait 76 paquets dont la description rapide contenait le mot « *meta-analysis* » sur le CRAN. Cela reflète l'intérêt de la communauté scientifique pour la recherche d'outils pour réaliser des méta-analyses, quel que soit le type de données. L'objet de ce manuscrit n'est pas de développer les avantages et inconvénients de chaque paquet, le travail ayant été amorcé dans la synthèse de Polanin, Hennessy, & Tanner-Smith (2017), qui ciblent préférentiellement les méta-analyses utilisant la taille de l'effet. Ils concluent sur l'absence de différences de performances notables entre les 11 paquets comparés. Ils recommandent notamment aux nouveaux utilisateurs l'utilisation de paquets de méta-analyse générauxet notamment le plus simple metafor.

Certains paquets sont spécifiques à un type de données, tel que le paquet MetaMSD, spécialisé dans le traitement des données de spectrométrie de masse (Ryu et Wendt, 2019), tandis que d'autres peuvent être utilisés quel que soit le domaine d'application (le plus connu étant metafor (Viechtbauer, 2019)). Deux paquets encore présents sur le CRAN ou Bioconductor sont spécifiques au traitement de données de puces à ADN : metaARRAY (disponible sur Bioconductor (Ghosh et Choi, 2008)), et metaMA (Marot, 2015). Les paquets metaDE et MAMA évoqués dans Polanin *et al.* (2017) ont été supprimés du CRAN (respectivement le 23/01/2018 et le 26/03/2017) et archivés, car des problèmes n'ont jamais été corrigés malgré plusieurs rappels. De même, le paquet MADAM a été archivé et supprimé du CRAN (le 01/07/2015) suite à un problème de contact avec le créateur. MetaMA est l'un des paquets pour combiner les valeurs p sur les données appariées ou non. De plus, il existe un paquet avec des fonctions similaires pour les données de RNAseq : metaRNAseq (Marot, Jaffrezic et Rau, 2013).

| Tableau 13 : Principaux paquet | s R existant pour la n | néta-analyse de doni | nées transcriptomique | es et leurs |
|---------------------------------|------------------------|----------------------|-----------------------|-------------|
| méthodes statistiques. Les paqu | ets en orange sont ce | eux qui ne sont plus | disponibles sur le CR | AN. |

| Paquets | Méthode de méta-analyse utilisée |
|-----------|---|
| MetaSD | Combinaison de valeurs p : Stouffer ou Pearson (normale ou chi carré) |
| Metafor | Taille d'effet : Fixé par l'utilisateur (fixe ou aléatoire) |
| MetaARRAY | Fusion directe |
| metaMA | Combinaison de valeurs p : Normale (inverse) |
| MAMA | Combinaison de rang |
| metaDE | Fixé par l'utilisateur : Combinaison de valeurs p, combinaison de rang, taille d'effet (FEM, REM) |
| MADAM | Combinaison de valeurs p : Fisher (chi carré) |

La méthode utilisée dans le paquet metaMA est la méthode inverse normale pondérée. La méthode inverse normale, ainsi appelée par Hedges et Olkin en 1985, a été introduite de manière indépendante par Stouffer et al.(1949) et Lipták et al.(1958), et définie comme la moyenne des valeurs p individuelles transformée en scores normaux et dont la définition mathématique est expliquée dans Marot, Foulley, Mayer, & Jaffrezic (2009). Elle présente l'avantage, contrairement à l'une des méthodes les plus utilisées qui est la méthode de Fisher, de traiter les gènes sous- et surexprimés ensembles, dans le sens où les valeurs p proches de zéro sont accumulées de la même manière que les valeurs p proches de l'unité. Elle est donc adaptée pour combiner des résultats de gènes différentiellement exprimés, quel que soit le sens de déviation par rapport à l'hypothèse nulle. Cette méthode a été testée sur des données de puces à ADN publiques et peut s'appliquer à des données en log ratio. Elle a été montrée par Loughin (2004) comme fonctionnant particulièrement bien dans les cas où les preuves contre l'hypothèse nulle sont réparties de manière égale entre les différentes études. La méthode normale a ainsi été montrée comme capable de rejeter l'hypothèse nulle (H₀:le gène n'est pas différentiellement exprimé au niveau de la méta-analyse) même lorsqu'une valeur p tend vers 1 (proche de 1) quand les autres restent petites (inférieur à 0.05), contrairement aux méthodes de type uniforme et maximum pour les valeurs p (Loughin, 2004).

Synthèse bibliographique – Méta-analyse



OBJECTIFS DE THESE

Il ressort de notre analyse bibliographique que des besoins persistent pour la qualification de l'adiposité de la carcasse et du muscle pour les filières de production bovine. Des méthodes sont disponibles pour les carcasses et pour les muscles piécés/viandes, et elles sont souvent spécifiques à une zone géographique, une race ou un tissu. La méthode prometteuse pour qualifier l'adiposité de la carcasse est la prédiction de la masse des tissus adipeux à partir du poids de carcasse chaude. Les équations de prédiction ont cependant été largement développées à partir de races précoces et restent à vérifier sur des races françaises (Al-Jammas, 2017). Les méthodes prometteuses pour qualifier l'adiposité musculaire sont la SPIR et l'analyse d'image avec une fiabilité importante (R > 0.80) lorsqu'elles sont appliquées à un muscle (Mourot et al., 2014 ; Meunier et al., non publié). Les rares méthodes pour qualifier l'adiposité corporelle ou musculaire du vivant de l'animal, sont extrêmement couteuses et non compatibles avec une utilisation terrain en particulier pour le bovin, ou avec une fiabilité qui reste à parfaire pour la SPIR. Aussi, dans un but de compréhension puis de manipulation par les conduites d'élevage de verrous moléculaires de l'adiposité des carcasses ou des muscles, de nombreuses études transcriptomiques ou protéomiques ont été appliquées à des muscles ou tissus adipeux bovins ces 20 dernières années.

Des variations de l'adiposité corporelle ou musculaire ont été induites par l'âge, la race, le sexe ou l'alimentation et ont été caractérisées par un transcriptome ou un protéome différentiel. En ce qui concerne les rares publications qui rapportent le protéome de tissus adipeux bovins, le nombre de protéines dont l'abondance varie est supérieur lorsque des sites anatomiques plutôt que des âges, races, ou régimes alimentaires sont comparés. Le site anatomique est un facteur de variation connu de la cellularité et de l'adiposité chez le bovin (Bonnet et al., 2010; Hausman et Canovas, 2017; Louveau, Perruchot, Bonnet et Gondret, 2016). **Aussi la comparaison des protéomes divergents selon les sites anatomiques est une source d'information potentielle sur les verrous moléculaires de la croissance des tissus adipeux corporels.**

D'autre part, nous avons vu que le persillé contribue aux qualités des viandes bovines. Des protéomes différentiels pour des bovins « viande » divergents par le taux de persillé étaient disponibles dans plusieurs études. Un premier effort de synthèse incluant les résultats de 2 publications (Baik et al., 2017) indiquait qu'une modification des métabolismes lipidique, glycolytique et contractile discriminait des bovins divergents par le persillé. Enfin, l'analyse bibliographique que nous avons faite révèle que le protéome musculaire divergent sur le persillé a été décrit essentiellement chez des races précoces et jamais chez des races tardives. Objectifs

Bien que les protéines soient une unité fonctionnelle de la cellule, le volume de données de transcriptomique en lien avec l'adiposité corporelle ou musculaire est largement supérieur au volume de données de protéomique. L'essentiel des données transcriptomiques relatives aux tissus adipeux et aux muscles sont disponibles dans des bases de données publiques et n'ont pas été exploitées à leur plus haut potentiel. Certains jeux de données disponibles n'ont pas été publiés et donc aucune connaissance ne leur est associée. D'un autre côté, un jeu de données peut avoir été valorisé dans de multiples publications ou plusieurs jeux de données analysés ensemble dans une publication unique. En outre, pour les données décrites dans les publications, il n'y a que peu de consensus entre les études, probablement du fait de la masse et de la diversité des informations disponibles ou de la diversité des facteurs de variation. Sur la base du recensement que nous avons réalisé, 23 jeux de données provenant de puces à ADN sont cependant disponibles pour tenter de révéler des verrous moléculaires de l'adiposité musculaire.

Mon projet de thèse s'inscrit dans un effort de l'équipe Biomarqueurs et plus largement de l'INRA, de satisfaire à la fois le consommateur et les filières de production de viande bovine, par l'identification et la compréhension des mécanismes métaboliques et physiologiques contrôlant la plasticité des tissus en lien avec la qualité des produits de ruminant. La finalité de ces connaissances est de concevoir des outils de phénotypage des qualités des produits. **Dans ce contexte, mon objectif de thèse était donc d'agréger et de ré-analyser des données publiques pour proposer des gènes ou protéines marqueurs de l'adiposité corporelle ou musculaire, de compléter les connaissances disponibles et d'identifier des manques de données.**

Pour répondre à cet objectif, ma thèse a été divisée en 2 parties. Une première partie porte sur l'agrégation et la complétion expérimentale de données protéomiques en lien avec l'adiposité corporelle et musculaire des bovins. Pour ce faire, j'ai compilé par fusion directe de données protéomiques les résultats en lien avec des différences d'adiposité entre sites anatomiques de tissus adipeux et en lien avec des variations de la teneur en lipides intramusculaires (Article 1). Cette synthèse ayant mis en évidence 50 biomarqueurs potentiels identifiés dans le muscle de races précoces, j'ai caractérisé par deux techniques (2DE et shotgun) le protéome musculaire de bovins Rouge des Prés divergents par leur teneur en lipides intramusculaires et le pourcentage de tissu adipeux dans la carcasse (Article 2).

La deuxième partie de mon travail a porté sur l'agrégation et l'analyse de données transcriptomiques publiques en lien avec l'adiposité musculaire des bovins. Pour ce faire, j'ai appliqué la procédure d'analyse décrite dans la synthèse bibliographique. Après avoir identifié

les données publiques disponibles, j'ai extrait des publications et des bases de données, les métadonnées descriptives des bovins, des échantillons, des méthodes, et les valeurs d'abondance. Sur la base de ces informations, j'ai catégorisé les données selon les facteurs de variation majeurs de l'adiposité musculaire. Les identifiants de gènes ont ensuite été mis à jour pour leur analyse par la méta-analyse. J'ai identifié les gènes différentiellement exprimés dans le muscle *Longissimus* chez des bovins divergents par l'âge, la race, le sexe et l'alimentation. Un premier article a pu être rédigé en prenant comme facteur de variation l'âge.

RESULTATS
Points clés de la synthèse bibliographique : article 1

9 L'ADIPOSITE CORPORELLE

Le tissu adipeux (TA) se développe dans différentes parties du corps du bovin telles que la cavité abdominale, la zone sous-cutanée ou encore autour (tissu adipeux intermusculaire) et dans (tissu adipeux intramusculaire) les muscles. Selon les sites anatomiques, les dynamiques de croissance sont différentes et sont à l'origine de fonctionnalités cellulaires et métaboliques spécifiques à un site anatomique (M. Bonnet et al., 2010; Hausman, Basu, Wei, Hausman et Dodson, 2014; Louveau et al., 2016). Ces fonctionnalités impactent la croissance des différents tissus adipeux, avec des conséquences sur les capacités de mobilisations des tissus internes chez le bovin vivant, ou sur la masse des tissus adipeux musculaires et sous-cutanés valorisés dans la carcasse bovine. Or il ressort de notre analyse bibliographique que peu de données rapportent le protéome des tissus adipeux et que les facteurs de variations ne sont pas facilement comparables. En outre, le site anatomique impacte plus le protéome des tissus adipeux que la nutrition par exemple. Aussi, nous avons rapporté et comparé le protéome des tissus adipeux en fonction des sites anatomiques chez des bovins et des caprins.



Figure 25: Comparaison de 3 publications portant sur les différences de protéome entre des dépôts adipeux chez un même animal. DA correspond à Différentiellement Abondantes.

Certaines annotations d'ontologie de gène (GO) sont spécifiques des tissus adipeux. Les GO spécifiques au tissu adipeux sous-cutané sont relatifs aux métabolismes du glucose et des lipides, à la prolifération cellulaire et à l'organisation de la matrice extracellulaire. Les GO spécifiques au tissu adipeux périrénal sont relatifs à la traduction des protéines, la modification des ARNr et ARNm et la réparation de l'ADN. Ces annotations spécifiques reflèteraient la moindre maturité métabolique du TA sous-cutané.

10 L'ADIPOSITE MUSCULAIRE

Le persillé (*marbling* en anglais) est le dépôt de tissu adipeux intramusculaire qui est visible dans la viande sous forme de fins filons blancs dans le muscle. La compréhension de ce dépôt adipeux particulier permettrait de développer des stratégies pour manipuler par les pratiques d'élevage le persillé bovin et ainsi produire des produits carnés en adéquation avec les attentes des filières de production bovines ou des consommateurs.

Pour comprendre et prédire la capacité de l'animal à déposer le tissu adipeux intramusculaire, 5 études ont rapporté le protéome différentiel entre des groupes de bovins divergents par des teneurs de persillé. Les méthodes de protéomique utilisées sont des analyses 2-DE et LC-MS/MS appliqués au muscle *longissimus lumborum* ou *longissimus dorsi* de bouvillons, majoritairement de races précoces. Dans ces études des variations de la teneur en lipides intramusculaires (persillé) ont été observées selon l'âge (Kim et al., 2009; Q. Zhang et al., 2010), la race (Keady et al., 2013) ou selon un score de persillage à 24 mois (Shen et al., 2012) et 30 mois (Mao et al., 2016).



Figure 26: Comparaison de 5 publications portant sur la différence de protéome entre des animaux différents par le persillé. DA correspond à Différentiellement Abondantes.

La fusion des données rapportées dans les cinq publications a permis d'identifier 50 protéines uniques dont l'abondance varie entre des groupes divergents par le persillé (Tableau 14). Parmi ces 50 protéines, seulement 9 sont présentes dans 2 publications et plus, avec parfois des sens de variations différents entre les études (Tableau 15). Ces résultats divergents pourraient résulter de différences de cellularité (rarement décrites) entre études. En effet, certaines voies métaboliques, en particulier les activités lipogéniques, sont supérieures dans les adipocytes de petits volumes comparativement aux adipocytes de gros volume (M. Bonnet et al., 2007; Romao et al., 2013).

Tableau 14 : Listes des 50 protéines uniques dont l'abondance varie selon la teneur en lipides intramusculaires. Cette liste a été produite à partir des 5 publications disponibles début 2017. Les protéines trouvées dans au moins 2 publications sont surlignées en gris.

| ACO2 | C4A | HP | OGDH | SH3BGR |
|--------|--------|-------|-------|---------|
| ACTA2 | CA2 | HSPB1 | PEBP1 | TNNI2 |
| ACTB | CAPZA2 | LDB3 | PGAM2 | TNNT1 |
| ACTN3 | CASQ1 | LDHB | PGK1 | TnT |
| ALDOA | CAST | MYH1 | PGM1 | TPI1 |
| AMPD1 | CCT8 | MYL1 | PRDX3 | UQCRC2 |
| APOE | COX4I1 | MYL3 | PRDX6 | VCL |
| ATP2A2 | COX6C | MYL6B | PYGM | VIM |
| ATP5H | FGA | MYLPF | RPLP0 | YWHAE |
| C2 | GRB2 | MYOZ3 | SDHA | ZSCAN31 |

Tableau 15: Relation entre la variation de l'abondance d'une protéine et la variation de teneur en lipides intramusculaires selon l'âge (Kim et al. 2009^1 ; Zhang et al. 2010^2), la race (Keady et al. 2013^3), ou le score de persillage à 24 mois (Shen et al. 2012^4) et 30 mois (Mao et al. 2016^5) chez les bovins

| Noms de gènes | Corrélation positive ou négative entre l'abondance des protéines et le persillé | | |
|---------------|--|--|--|
| CA2 | Négative ^{2,4,5} | | |
| MYL1 | Négative ^{3,4} et positive ¹ | | |
| HSPB1 | Négative ⁵ et positive ^{2,3} | | |
| TPI1 | Négative ^{1,5} et positive ⁴ | | |
| CAPZA2 | Positive ^{3,4} | | |
| MYLPF | Négative ^{3,5} | | |
| MYL3 | Négative ^{2,5} | | |

Proteomics Research in the Adipose Tissue

Fabrizio Ceciliani, Cristina Lecchi, Jeanne Bazile, and Muriel Bonnet

Abstract Adipose tissue is no longer considered only as a passive fuel reservoir, but a major endocrine organ, distributed in different depots throughout the body and actively involved in complex regulatory processes including appetite, energy expenditure, body weight, inflammation and reproduction. Proteomics has emerged as a valuable technique to characterize both cellular and secreted proteomes from adipose tissues. Adipose tissue is of major importance in farm animals: in dairy animals, it regulates energy metabolism as well as other functions. In cows and pigs, adipose tissue depots' distribution is of fundamental importance for the quality of carcasses. In this chapter, we provide a general overview of adipose tissue functions and its importance in farm animals and summarize the state of art on farm animal adipose tissue proteomics in cattle and pigs but also in chicken and in farmed fish.

Keywords Adipose tissue • Proteomics • Farm animals • Animal physiology

1 Introduction

Adipose tissue is a complex structure composed of several types of cells, including adipocytes and pre-adipocytes, macrophages, endothelial cells, fibroblasts and leucocytes. Adipose tissue historical function is to store the energy in excess, this role being already established for many years. This task is carried out by conserving the heat of the body and controlling the mobilization of lipids (Sethi and Vidal-Puig 2007). When energy is in surplus, it is efficiently deposited as neutral triglycerides in adipose tissue. When energy is required, adipocyte triglycerides are broken down into glycerol and fatty acids. The released glycerol and fatty acids are then

F. Ceciliani (🖂) • C. Lecchi

J. Bazile • M. Bonnet INRA, VetAgro Sup, UMR Herbivores, Université Clermont Auvergne, Saint-Genès-Champanelle, France

© Springer International Publishing AG 2018

Department of Veterinary Medicine, Università degli studi di Milano, Milano, Italy e-mail: fabrizio.ceciliani@unimi.it

A.M. de Almeida et al. (eds.), *Proteomics in Domestic Animals: from Farm to Systems Biology*, https://doi.org/10.1007/978-3-319-69682-9_12

transported in the blood and subsequently uptaken by the muscles, liver and other organs, orchestrating lipid distribution and whole-body energy balance (Frayn 2002). In farm animal physiology, most of the studies on adipose tissues are focused on specific mechanisms related to animal production, aiming to understand the biochemical and molecular background underlying the mobilization of lipids in lactating animals to fulfil the energy requirements of milk production or the deposition of fat in specific fat depots, to improve the quality of meat. A further role as a major player during systemic metabolic regulation has been increasingly recognized (Luo and Liu 2016). With the discovery of leptin in 1994 (Zhang et al. 1994), adipose tissue was no longer considered as a passive fuel reservoir, but a major endocrine organ, distributed in different depots throughout the body and actively involved in complex regulatory processes including appetite, energy expenditure, body weight, inflammation and reproduction by synthesizing and secreting messenger molecules which are now collectively referred to as "adipokines". Each of these depots plays an essential role in energy homeostasis as well as in endocrine regulation, at both local and systemic levels. The excessive accumulation of fat in adipose tissue causes obesity, a chronic disease that in turn drives obesity-associated metabolic disorders such as diabetes. Although not directly related to obesity, the dysregulation of adipose tissue metabolism represents an issue for farm animals as well. Adipose tissue is involved in the fine-tuning of energy turnover in dairy cows or in carcass and meat qualities in cows and pigs. Therefore, a deeper understanding of the adipose machinery regulation in rearing species can provide tools to overcome metabolic stress-related diseases and improve the quality of carcass and meat.

Among omics disciplines, transcriptomics has provided important advances in understanding the functions of adipose tissue. A major weak point of transcriptomics is that it does not provide any hint about the effective abundance of the proteins, the correlation between the abundances of mRNA and of protein being poor (Griffin et al. 2002). Integration of data derived from gene expression analyses with those from protein expression studies is therefore mandatory to draw a complete and reliable picture of the functions and the activities of adipose tissue. The proteomics of adipose tissue was recently reviewed (Sauerwein et al. 2014). In the present chapter, we will try to provide the reader with a general overview of adipose tissues and summarize the recent advances on adipose tissue proteomics, focusing on farm animals such as cattle and pigs but also chicken and fish.

2 Adipose Tissue: A Bird's-Eye View with a Proteomic Perspective

2.1 The Main Structure and Function of Adipose Tissue: White, Brown and Beige Adipose Tissue

Two main types of adipose tissue exist in mammals, based on differences in morphology, location and functions: white adipose tissue (WAT) and brown

adipose tissue (BAT) (Lizcano and Vargas 2016). WAT is the preferential site to store energy in the form of triacylglycerols during excessive energy disposal and to restore it during fasting periods. Beside this undisputed role, WAT is also involved in systemic metabolic regulation and inflammation. The main role of BAT is to dissipate chemical energy as heat via high levels of uncoupling protein 1 (UCP1), regulating hypothermia by metabolizing glucose and lipids to produce heat participating to non-shivering thermogenesis (Lizcano and Vargas 2016; Louveau et al. 2016).

Recent studies have described another type of thermogenic adipocytes known as beige/brite (brown in white) adipocytes. Beige adipocytes share morphological and functional similarities with brown ones. In BAT, cold stress or β 3-adrenoceptor agonists that mimic cold stress (Barbatelli et al. 2010) stimulate UCP1 expression. Therefore, both brown and beige fat fulfil thermogenic roles. In farm animals, the presence of beige or brown-like adipocytes has been suggested in white adipose depots of fattening cattle (Asano et al. 2013) and sheep (Pope et al. 2014).

2.2 Different Depots with Different Proteomes: Comparative Analysis of Subcutaneous and Visceral White Adipose Tissue Proteomics

Adipose tissue develops in various anatomical sites, including the abdominal cavity, the subcutaneous districts and the musculature. They are known to display a different dynamic of growth, which is at the origin of cellular and metabolic features (Bonnet et al. 2010; Hausman et al. 2014; Louveau et al. 2016). Thus, adipose tissue depots respond differentially to rearing practices, which may have implications on the metabolic turnover of lipids and nutrients at the body level for metabolic adaptations, or on carcass qualities and valorization, as visceral fat is discarded, while subcutaneous fat is partly consumed with the muscles. To date, few studies have attempted to investigate the proteome differences among fat depots in ruminants.

A study relying on 2-DGE paired with sequencing mass spectrometry investigated the proteins in undifferentiated and differentiated preadipocytes collected from bovine omental, subcutaneous and intramuscular adipose depots, identifying differentially expressed intracellular proteins during adipogenic conversion (Rajesh et al. 2010). A total of 65 proteins were found to be differentially abundant across the three depots. The preadipocyte differentiation induced a downregulation of many structural proteins, whereas proteins associated with lipid metabolism and metabolic activity, including ubiquinol-cytochrome-c reductase complex core protein I (UQCRC1), ATP synthase D chain, superoxide dismutase (SOD), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), sulfotransferase 1A1 (SULT1A1), carnitine *O*-palmitoyltransferase 2 (CPT2) and heat shock protein beta 1 (HSPB1), were found to be upregulated across the three depots. Among these proteins, five (tropomyosin alpha-4 chain, rho GDP-dissociation inhibitor 1, purine nucleoside phosphorylase, transgelin, non-muscle caldesmon), one (annexin A1) and four (glyceraldehyde-3-phosphate dehydrogenase, prohibitin, voltage-dependent anion-selective channel protein 2, voltage-dependent anion-selective channel protein 1) proteins were specifically identified during the adipogenic process of omental, subcutaneous and intramuscular cells, respectively. Using a label-free quantification LC-MS/MS, 682 proteins were identified both in subcutaneous and visceral adipose tissues of 15.5-month-old British continental steers regardless of the diet. The abundance of 51% of these proteins was modified depending on the anatomical site. Of these, 240 were known proteins and were assigned to the following top categories of molecular and cellular functions: lipid metabolism, energy production, small molecule biochemistry, post-translational modification, cellular assembly and organization, cell morphology, protein synthesis and cellular function and maintenance (Romao et al. 2013). Notably, proteins related to oxidation of lipids and fatty acids and synthesis of lipids were more abundant in visceral than subcutaneous adipose tissue, in agreement with the repeatedly observed greater metabolic activity in visceral than in subcutaneous adipose tissues. However, any of the identified proteins were reported to be specifically present in one adipose tissue depot. To date, the sole study in ruminants that reports on protein signatures that are specific to the depot location was conducted in the goat (Restelli et al. 2014), in which a LC-MS/MS approach has identified 833 proteins in at least one of the four adipose tissue deposits investigated in 1-month-old goat kids. The tissue coverage was 471 proteins identified in subcutaneous adipose tissue sampled from the base tail, 480 proteins in subcutaneous adipose tissue sampled from the sternum, 587 proteins in omental tissue and 654 proteins in perirenal adipose tissue. These proteins were assigned to four major functional categories, based on similarity of functions, namely, (1) metabolic processes (62.1% of proteins); (2) cell adhesion cytoskeleton, intramuscular transport and membrane integrity (22.2% of proteins); (3) toxic response and folding (9.5% of proteins); and (4) proteins involved in immune and inflammatory response (6.1% of proteins). Of these 39, 30, 72 and 27 proteins were identified in only one adipose depot, the subcutaneous sampled from the base tail and the sternum as well as perirenal and omental depots, respectively. For the purpose of this review, we focused on the lists of proteins specifically identified in subcutaneous, from the base tail and of visceral perirenal adipose tissue, which are the most commonly analysed, in order to identify proteins or pathways related to the depot specificity of fat accumulation. Thus, lists of proteins specifically found in subcutaneous or perirenal adipose tissues were subjected to annotations according to the gene ontology (GO) with ProteINSIDE (Kaspric et al. 2015) in order to provide the biological processes in which they are involved. The 59 (over 72) perirenal and the 35 (over 39) subcutaneous proteins were annotated by 193 and 129 significantly enriched GO terms, respectively. Of these only 16 GO terms were found in common but annotating different proteins with a different *p*-value of enrichment between lists of subcutaneous and perirenal proteins (Fig. 1). This may reveal pathways important for adipose tissue growth in 1-month-old goat kids, however, involving proteins that



and 0.05, respectively. In brackets are the gene names of the proteins annotated by GO terms

differ depending on the anatomical site of adipose tissue. The top ten of the most enriched GO terms that were specifically found for the subcutaneous proteins were related to glucose or lipid metabolism, cell proliferation and extracellular matrix organization, in line with a tissue growth both by hyperplasia and by hypertrophy (Fig. 2). For the perirenal proteins, the top ten of the most enriched GO terms were related to protein translation, rRNA and mRNA processing and DNA repair (Fig. 3). These divergent annotations may reflect the lower maturity of subcutaneous as compared to perirenal adipose tissue, in line with the known depot-specific



Fig. 2 Relations between the most enriched GO terms were visualized as networks provided by ProteINSIDE. The network links the most enriched GO terms (*p*-value < 0.05) in the biological process category for the list of 72 proteins (59 annotated) from the perirenal adipose tissue (1) and for the list of 39 (35 annotated) proteins from the base tail subcutaneous adipose tissue (2) in goat. The degree of colour saturation is related to the number of proteins annotated by a GO (dark and clear for high and low numbers, respectively)



Fig. 3 Distribution and overlap of proteins identified in five proteomic analysis (Kim et al. 2009; Zhang et al. 2010; Shen et al. 2012; Keady et al. 2013; Mao et al. 2016) carried out to provide the molecular basis of marbling and to identify biomarkers. Venn diagram was proceeded with http://bioinformatics.psb.ugent.be/webtools/Venn/

growth patterns reported during the perinatal period in ruminants (Bonnet et al. 2010). Future proteomic studies might contribute to identify depot-specific proteins in order to increment available data and provide the molecular basis of depot specificity of fat accumulation for a robust data mining of the pathways or molecular functions involved.

2.3 Proteomic Analysis of Brown Adipose Tissue

Larger mammals such as ruminants are fully developed and able to thermoregulate at birth, thanks to the presence of BAT that disappears after birth and is replaced by WAT. A better knowledge of the molecular features of BAT versus WAT may help in the understanding of the balance between brown and white fat cells and is a prerequisite for a better control of thermogenesis and survival in perinatal period, as well as energy efficiency during the productive life of the ruminants. To the best of our knowledge, only one study has attempted to link the proteome to chemical composition, cellularity, histology, enzyme activities and gene expression during a time-course analysis of the ontogeny of perirenal adipose tissue in bovine foetuses (Taga et al. 2012). Between 110 and 260 days postconception (dpc, 38 and 90% of gestation length, respectively), the increase in the weight of perirenal adipose tissue resulted from an increase in the adipocyte volume and mainly number and was accompanied by changes in the abundance of 128 proteins among the 143 proteins identified. From these data, an unexpectedly high abundance of the α -subunit of ATP synthase, a member of complex V (ATP synthase) of the respiratory chain which is normally bypassed in BAT, and of aldehyde dehydrogenases ALDH2 and ALDH9A1, has been observed at 180 dpc when compared to other foetal ages. However, these proteins were proposed as hallmark WAT in mouse (Forner et al. 2009). From 210 dpc, an increase in the mRNA abundances of UCP1 and DIO2, which are hallmarks of brown adipocytes, was observed. These results, associated with the presence of numerous unilocular white adipocytes and few brown multilocular adipocytes from 180 dpc, show that foetal ATs up to 260 dpc have molecular and cellular characteristics of WAT in addition to those of BAT (Taga et al. 2012). These results challenged the commonly accepted concept that foetal perirenal AT in cow is brown and highlighted that perirenal AT is a heterogeneous tissue made of brown and white adipocytes up to 260 dpc in bovine foetuses. This study, together with the global gene expression profiling of brown/beige to white adipose tissue during the first month of life in lambs (Basse et al. 2015), provides a useful resource to identify the molecular features of perinatal brown/beige adipose tissue relatively to those of white adipose tissues presented above.

3 Adipose Tissue in Farm Animals: Endocrinology and Immunology

With the discovery of leptin in 1994 (Zhang et al. 1994), adipose tissue has been identified as an endocrine gland secreting messenger molecules that affect a wide variety of bodily functions. From then on, numerous hormones, chemokines, cytokines as well as acute-phase proteins were discovered as adipose products and are now collectively termed as adipokines. Adipokines are involved in the dynamic control of energy metabolism. Being produced by adipose tissue, adipokines are fundamental to deliver to other tissues important information about the nutrient status of the organism, in particular to those responsible for controlling energy intake and expenditure, such as appetite regulation, energy expenditure, insulin sensitivity, glucose metabolism and fatty acid oxidation. Figure 4 presents a short list of adipokines and their main physiological roles. The human adipose tissue secretome determined by proteomic techniques was released in 2007 (Alvarez-Llamas et al. 2007). Products from all cell types present in adipose tissue were included in this first report. In fact, adipose tissue contains not only adipocytes but other cell types as well. A stromal vascular fraction of cells including preadipocytes, fibroblasts, vascular endothelial cells and a variety of immune cells (e.g. macrophages) are also included in adipose tissue depots.



Fig. 4 The complexity of AT and physiological functions of adipokines. Beside adipocytes, adipose tissue also contains preadipocytes; fibroblasts and connective tissues; vascular endothelial cells, providing the capillary network; myocytes; and muscle cells and macrophages that increase when adipose tissue becomes inflamed. In this figure, some of the proteins that are produced by AT, and that can be found in the circulation, the so-called adipokines, are also presented. Some of their main activities are also briefly summarized. *FGF21* fibroblast growth factor 21, *IL-6* interleukin 6, *MCP1* monocyte chemoattractant protein 1, *PAI1* plasminogen activator inhibitor 1, *TNF-* α tumour necrosis factor alpha and *PEDF* pigment epithelium-derived factor. We thank Mrs Helen Arino for drawing the figure

Proteomics analysis carried out on primary human adipocyte cell culture identified 347 proteins of which 263 were predicted to be secreted (Lehr et al. 2012). The list of these which are regarded as adipokines includes not only proteins whose abundance is restricted to adipose tissue but also molecules taken up by adipose tissues that are known to be secreted from other tissues, such as cytokines and acute-phase proteins. Most of adipokines that were initially discovered and considered as adipose specific were then found to be secreted by other tissues such as skeletal muscle and are now classified as myokines. Both adipokines and myokines play major roles in exchanging information between skeletal muscle fibres and adipocytes in the framework of an endocrine cross talk (Trayhurn et al. 2011). Recently, a computational prediction of the large-scale secretome of adipose tissues and muscles in ruminant was achieved by applying algorithms to 24 publications reporting transcriptomics or proteomics results from bovine, ovine and caprine species and eight transcriptomics dataset series. In this way, 1749 proteins were proposed as secreted both by adipose tissues and muscles, and 188 and 357 proteins were predicted to be secreted either by adipose tissues or muscles, respectively (Bonnet et al. 2016).

As a consequence of the cell diversity in adipose tissue, during human obesity, macrophages forming a part of the stromal vascular fraction infiltrate adipose tissue

| montais of age in boy | |
|-----------------------|--|
| Gene names | Positive or negative correlations between protein abundance and marbling |
| CA2 | Negative ^{2,4,5} |
| MYL1 | Negative ^{3,4} and positive ¹ |
| HSPB1 | Negative ⁵ and positive ^{2,3} |
| TPI1 | Negative ^{1,5} and positive ⁴ |
| CAPZA2 | Positive ^{3,4} |
| MYLPF | Negative ^{3,5} |
| MYL3 | Negative ^{2,5} |
| PYGM | Negative ⁵ and positive ³ |
| PGK1 | Negative ⁵ and positive ⁴ |

Table 1 Proteins identified as related to intramuscular fat deposition in at least two studies from five, in which strong differences in marbling resulted from age (Kim et al. 2009^1 ; Zhang et al. 2010^2), breed (Keady et al. 2013^3) or marbling score at 24 (Shen et al. 2012^4) or 30 (Mao et al. 2016^5) months of age in bovine

Superscripts refer to the publications as defined in the table caption

in increasing numbers, providing important contribution to the secretory function of adipose tissue, in particular for inflammatory cytokines, such as TNF- α and IL-6 (Weisber et al. 2003). The increase in circulating levels of pro-inflammatory macrophage-derived factors during obesity induces a chronic low-grade inflammatory state leading to the development of insulin resistance and diabetes (Xu et al. 2003). It must be said that immune cells were hardly detectable in both visceral and subcutaneous adipose tissue in dairy cattle. Therefore, macrophages might not be involved in the immunity and metabolism of adipose tissue in nonobese lactating animals (Akter et al. 2012). In the previously described study on goat adipose tissue proteomics (Restelli et al. 2014), it was demonstrated that goat adipose tissue contains 51 proteins that are related to inflammatory and immune responses. The list of these proteins includes acute-phase proteins, such as C-reactive protein, ceruloplasmin and alpha-1-acid glycoprotein and galactoside-binding proteins (LGALS), such as beta-galactoside-binding lectin precursor and galactoside 3. LGALS act as agonists of platelet activation and are pro-apoptotic for immune cells, activate and increase the adhesion of neutrophil and are chemotactic for monocytes (Table 1). Whether these proteins are produced within adipose tissue or are present in adipose tissue because of their uptake from the blood remains to be investigated in ruminants.

4 When Things Turn Wrong: Adipose Tissue in Dairy Animals and Metabolic Imbalance Syndrome

Hepatic lipidosis (also known as "fatty liver disease" or "fat cow syndrome") is a common production problem of dairy cows occurring during the critical physiologic transition from pregnancy to lactation (Bobe et al. 2004; Hammon et al.

2009). During recent decades, the genetic selection carried out on dairy cows was directed toward an increase of milk production. The result was positive, since milk production reached performance peaks unthinkable even a few years ago. This major gain did not come without serious consequences for cow health. The excessive demand for nutrients frequently ends up in a severe energetic deficit at the beginning of lactation (Mulligan and Doherty 2008; Hammon et al. 2009). The severity of this effect is increased by the need of nutrients for the foetus that is growing exponentially and the fact that feed intake does not equally increase. Animals affected by this major energy deficiency try to counteract by developing insulin resistance (IR) to spare glucose for the foetus (Bauman and Currie 1980) and rapidly mobilizing fat depots, in the attempt of providing nonesterified fatty acids (NEFA) as an energy source. Triacylglycerols are released from hepatocytes as part of lipoproteins, mostly very low-density lipoproteins (VLDL). In ruminants, the secretion of VLDL from the liver is limited as compared with other species. Consequently, the hepatic uptake of NEFA and storage in the form of triacylglycerols exceed their elimination. The storage of excess lipids in hepatocytes leads to liver damage and depressed liver functions (Geelen and Wensing 2006). A morphologic change in liver tissue occurs, characterized by the accumulation of lipid vacuoles within hepatocytes. This metabolic change is called hepatic lipidosis or liver fatty change (Cebra et al. 1997; Imhalsy et al. 2014). Considering this dual role of regulating energy storage by storing and releasing fatty acids, and acting as a major endocrine capable of influencing metabolism by secreting and regulating hormones and adipokines (Kershaw and Flier 2004), adipose tissue metabolism plays an essential role in the development of metabolism syndrome in transition dairy cows. Adipose tissue reacts to the increase of energy demand by increasing lipolysis and by regulating the development of major metabolic changes including, among the other, insulin resistance or sensitivity (Bell and Bauman 1997; Rabe et al. 2008; Loor et al. 2013).

A recent study determined the proteome of transition cow adipose tissue focusing on the relationship between insulin sensitivity (IS) and resistance (IR) of subcutaneous adipose tissue (Zachut 2015). The results showed that a number of 143 proteins out of 586 were differentially abundant in prepartum versus postpartum adipose tissue. The lipid metabolism-related functions that were significantly changed postpartum compared with prepartum include fatty acid metabolism, the esterification of lipids and oxidation of fatty acids. The proteins whose abundance was decreased in postpartum included fatty acid synthase, complement C3, annexin A1 and acyl-CoA desaturase. Finally, the study also addressed proteomic differences between IR and IS adipose tissues. Out of 586 proteins, 111 were differentially abundant between IS and IR cows. Most of them (a number of 106) were increased in IR versus IS adipose, whereas only five were decreased. As expected, the most relevant pathways differentially activated between IR and IS adipose tissue included energy-related pathways, such as gluconeogenesis and glycolysis, 14-3-3-mediated signalling, TCA cycle and ERK/MAPK signalling. The most relevant function in IR as compared to IS adipose tissue was inflammatory response, such as leukocyte migration and proliferation of T lymphocytes, confirming the relationship between adipose tissue IR and proteins related to inflammation, and organization of actin cytoskeleton.

5 The Marbling Issue: Adipose Tissue in Beef

Understanding muscular adipogenesis and identifying a protein profile associated with intramuscular fat deposition also termed marbling are prerequisites to develop strategies to manipulate marbling in cow and pig in an effort to produce even healthier and tastier meat products for consumers. In order to provide the molecular basis of marbling and to identify biomarkers to predict the animal's ability to deposit intramuscular fat, proteomics of fat accumulation was assayed in the bovine species in five proteomic investigations mainly through 2DE followed by MS/MS analysis. The proteome of *longissimus lumborum* or *longissimus dorsi* was explored in steers, mainly of Asian breeds, showing strong differences in marbling as the result of age (Zhang et al. 2010; Kim et al. 2009), breed (Keady et al. 2013) or marbling score at 24 (Shen et al. 2012) or 30 (Mao et al. 2016) months of age that are among the major drivers of individual variations in the intramuscular fat percentage (Shingfield et al. 2013).

By merging results from these five studies, 50 unique proteins were proposed to be involved in the deposition of intramuscular fat content that are informative of the molecular basis and of the major molecular pathways involved. Among them four (Zhang et al. 2010) and seven (Kim et al. 2009) proteins were found to be related to a high intramuscular fat deposition induced by an increase in age of Korean steers. These proteins were heat shock protein beta 1 (HSPB1, upregulated), ATP synthase D chain mitochondria (ATP5H), carbonic anhydrase II (CA2), myosin light chain 3, slow-twitch muscle [MYL3, all downregulated (Zhang et al. 2010)], as well as myosin light chain 1, slow-twitch muscle A isoform (MYL1), thioredoxindependent peroxide reductase (PRDX3, both upregulated), actin, aortic smooth muscle (ACTA2), actin, cytoplasmic 1 (ACTB), succinate dehydrogenase [ubiquinone] flavoprotein subunit (SDHA), triosephosphate isomerase (TPI1) and zinc finger protein 323 [ZSCAN31, downregulated (Kim et al. 2009)]. These proteins are mainly involved in glycolysis/gluconeogenesis (TPI1), the oxidation of glucose and fatty acids (ATP5H) and the regulation of muscle contraction (MYL3, MYL1, ACTA2). Additionally, the most enriched GO terms provided by ProteINSIDE (Kaspric et al. 2015) for these 11 proteins are "muscle filament sliding" (MYL1 MYL3), "muscle contraction" (MYL1 ACTA2), "vascular endothelial growth factor receptor signalling pathway" (ACTB HSPB1) and "negative regulation of apoptotic process" (HSPB1 PRDX3).

Differential abundances of 7 and 28 proteins were found to be related to divergent marbling scores, in Xiangxi yellow \times Angus cattle steers (Mao et al. 2016) and in Hanwoo cattles (Shen et al. 2012), respectively, three of them being reported in both breeds. The abundance of carbonic anhydrase II (CA2) was reported to be negatively related to marbling score in both breeds. However,

negative relations were reported between the abundances of triosephosphate isomerase (TPI1), phosphoglycerate kinase 1(PGK1) and the intramuscular fat content in the Xiangxi yellow \times Angus cattle steers, while a positive relation was reported in Hanwoo cattle.

Differential abundances of 17 proteins were found to be related to divergent intramuscular fat content in the muscle of Belgian Blue \times Holstein Friesian compared to Aberdeen Angus \times Holstein Friesian (Keady et al. 2013). The top canonical pathways identified by the authors were glycolysis/gluconeogenesis (glycogen phosphorylase (PYGM), phosphoglycerate mutase 2 (PGAM2) and aldolase A (ALDOA) as greater in abundance in highly marbled Aberdeen Angus), the citric cycle (aconitase 2 (ACO2) and 2-oxoglutarate dehydrogenase (OGDH) as greater in abundance in highly marbled Aberdeen Angus), the protein kinase A signalling pathway (myosin light chain 1 (MYL1), myosin light chain, phosphorylatable (MYLPF), PYGM and troponin I (TNNI2) differing in abundance across breed), and the pentose phosphate pathway [with ALDOA greater and phosphoglucomutase (PGM1) decreased in abundance in Aberdeen Angus compared to Aberdeen Angus].

Altogether these data highlight that the main pathways related to intramuscular fat deposition in beef are metabolic pathways related to glucose, oxidative pathways, the molecular features of muscle fibres (and thus of muscle type) and pathways related to apoptosis. Some of them could be related to previous observations that glucose rather than acetate is a major precursor for lipogenesis, and hence fat accumulation, within intramuscular adipocytes (Smith and Crouse 1984) or to the repeatedly observed higher marbling in skeletal muscle of type I than II (Bonnet et al. 2007).

Of these 50 unique proteins, only 9 were related to intramuscular fat deposition in at least two publications and thus could be potential robust biomarkers to predict meat marbling (Fig. 3 and Table 1). The proteins HSPB1, MYL1, CA2 and TPI1 were declared as differentially abundant according to the intramuscular fat deposition in three studies. Of these, CA2 was always reported to be lower in abundance in muscle with high intramuscular fat content. The proteins CAPZA2, MYLPF, PYGM, MYL3 and PGK1 were declared differentially abundant according to the intramuscular fat deposition in two publications. Among them, CAPZA was always reported to be higher in abundance in muscle with high intramuscular fat content, while MYLPF and MYL3 were reported to be lower (Table 1). The abundance of HSPB1, MYL1, TPI1, PYGM and PGK1 was reported to have a positive or negative relation with marbling depending on studies. The apparent inconsistencies of these results may be explained by an increase in some pathways related to lipogenic capabilities in adipocytes of small volume that are then decreased in large adipocytes, as already observed for proteins involved in lipid synthesis (Bonnet et al. 2007; Romao et al. 2014). A variation in the abundances of HSPB1, MYL1, TPI1, PYGM and PGK1 may partially explain the result of differences depending on breed, age, nutrition and not only of marbling. To sum up, CA2, CAPZA, MYLPF and MYL3 may be proposed as robust biomarkers of marbling in beef since their abundances were related to intramuscular fat deposition regardless of breed, age or level of marbling.

This literature review lastly shows that proteins related to marbling (Zhang et al. 2010; Kim et al. 2009; Keady et al. 2013; Shen et al. 2012; Mao et al. 2016) are different than those related to back fat thicknesses (Charolaise \times Red Angus and Hereford \times Angus; Zhao et al. 2010) in steers.

6 Adipose Tissue in Pigs

The pig is a major source of meat for human consumption. Knowledge of the biochemical mechanisms and understanding the biological significance of adipose tissue development in pigs are important for optimal growth efficiency and meat quality. A specific issue of porcine production is the processed meat products. The dynamic of lean and fat growth is of fundamental importance to improve the production of differentiated end products through breeding and feeding strategies. As an example, optimal adipogenesis requirement is different if the final product is fresh versus processed meat products, such as the production of regional and traditionally cured and dried high-value products, like Parma and Serrano ham (Candek-Potokar and Skrlep 2012). Meanwhile, with the background that swine physiology, genomics and nutritional requirements are very similar to that of humans, the pig has become an increasingly important animal model for human metabolic diseases and obesity (De Almeida and Bendixen 2012; Ceciliani et al. 2014). Despite these various interests, proteomic experiments to understand adipogenesis at the protein level in pig are still very limited, as compared to the information made available from genome-wide association, targeted candidate genes and large-scale transcriptomic studies (Cánovas et al. 2010; Ramayo-Caldas et al. 2012; Corominas et al. 2013; Puig-Oliveras et al. 2014; Xing et al. 2015, 2016; Shen et al. 2015; Ros-Freixedes et al. 2016; Zhang et al. 2016). Proteome research has mainly been focused on mapping the muscle proteome with postmortem modifications driving the transformation of muscle to meat, such as meat tenderness, postmortem proteolysis (Pioselli et al. 2011; Mora et al. 2015; Gallego et al. 2016), phosphoproteomics (Huang et al. 2014), water holding capacity (Di Luca et al. 2013, 2016) and meat colour traits (Lomiwes et al. 2014). Very few studies have been carried out on the pig adipose tissue proteome. The available proteomics were carried out in muscle with the aim to identify protein and pathways related to intramuscular fat deposition and in subcutaneous adipose tissue to identify pathways related to body adiposity.

Although intramuscular fat content is an important determinant of meat quality, the information related to deposition of intramuscular fat at protein and proteome levels is scarce. Differences between high and low ability for fat deposition in longissimus dorsi muscle from commercial crosses originating from Pietrain and an industrial cross originating from Duroc, Hampshire and Large White founders were identified in comparative study including both transcriptome and proteome profiles (Liu et al. 2009). Differences for marbling were correlated to transcriptome and proteome profiles, and 40 muscular genes were identified as differently expressed between high-fat and low-fat groups, either at the mRNA level (29) or encoding proteins (12). Among them, only GSTP1 was found differently abundant at both the mRNA and protein levels. Gene ontology analysis indicated that differentially expressed genes were involved in metabolic processing, cell communication, binding and response to stimulus. The results of this study suggest that interindividual variability in intramuscular fat content might arise essentially from differences in early muscular adipogenesis that may have modified either the volume or the number of intramuscular adipocytes. A recent investigation, pairing transcriptomics with proteomics, identified several genes and proteins involved in fatty acid metabolism and intramuscular fat deposition (Yang et al. 2016). A total of 23 differentially expressed proteins were identified, several of which were potentially associated with fatty acid metabolism. In particular, adipocyte fatty acid-binding protein A, alpha-enolase isoform X1 and beta-enolase isoform X1 were more abundantly found in Chinese indigenous Shaziling pig, a fat-type line with high intramuscular fat (IMF), as compared with the Yorkshire breed, which has a leaner meat ratio. By identifying differences in breed-related protein and transcript abundance patterns between the two breeds, these data provide insights into the mechanisms of growth and development of porcine skeletal muscle and how it might influence the IMF deposition.

The main adipose tissue studied in pig was the subcutaneous. Indeed, as compared to cattle which has a major deposition of fat within visceral and intermuscular adipose tissues, pigs deposit fat mostly within subcutaneous adipose tissue. An excessive growth of subcutaneous adipose tissue decreases the gain production of pig. A characterization of subcutaneous adipose tissue proteomes of young piglets was carried out focusing on the metabolic control and acute-phase response associated with adipogenesis in lean (Large White) as compared to fat (Basque) breeds (Gondret et al. 2012). Mass spectrometry and MS/MS analyses identified 65 proteins, 57 of which were significantly different between subcutaneous adipose tissues from Basque and Large White pigs, and 12 others were expressed with different abundance. The authors demonstrated that several metabolic pathways are differentially expressed in the lean versus fat breeds including, most significantly, the pentose-phosphate pathway, with aldolase C, glucose-6-phosphate dehydrogenase and ribose 5-phosphate isomerase A; the citrate cycle with pyruvate carboxvlase, dihydrolipovl dehydrogenase and aconitase; and the pyruvate metabolism with aldolase C and dihydrolipoyl dehydrogenase; again, malic enzyme; and pyruvate carboxylase proteins. The most significant biological function affected includes lipid metabolism, which was increased in Basque breed as compared to Large White lean one. Carboxylesterase 1 was more abundant in adipose tissue of Basque breed, suggesting an increased activity of triacylglycerol synthesis and degradation aimed to protect cells against fat overload. Adipose tissue of Basque piglets also showed an increased abundance of inflammatory-related proteins, suggesting a low-grade inflammation in "fat" pigs, probably induced by an increased oxidative stress related to adipocyte differentiation. This effect closely resembles human obesity and inflammatory metabolic syndrome status. The increased abundance of selenium-binding protein, which has ROS scavenging properties, also confirms the need of quenching ROS aiming to protect tissues from oxidative damages. To understand the impact of maternal nutrition on prenatal and offspring metabolism and body adiposity, the proteomes of subcutaneous adipose tissues in piglets born from sows fed on either low, high or normal protein diets were determined. The 2D proteome profiles between the three groups were different (Sarr et al. 2010, 2012). In details, proteins related to fatty acid metabolism and lipid transport were upregulated in piglets born from sows fed on low protein diets. Analogue transcriptome studies have confirmed that foetal programming indeed affects adiposity of pigs (Oster et al. 2011, 2012). Besides increasing our knowledge about the relationship between maternal nutrition and offspring adipose tissue development in human nutrition, these two studies also provide important information that might be implemented in improving rearing and possibly driving the practices affecting the partitioning of fat between subcutaneous and intramuscular adipose tissues.

7 Adipose Tissue in Other Species: Fish and Poultry

7.1 Adipose Tissue Proteomics in Broiler Production

Adipose tissue is regarded as a negative trait in poultry science, although it must be said that genetic pressure for rapid growth had also an increase of fat deposition as side effects (Nones et al. 2006). Proteomic analysis was carried out to obtain insights into the molecular basis of fat deposition in broiler combining 2D electrophoresis with identification by MALDI-TOF (Wang et al. 2009). Approximately 1000 protein spots were identified in adipose tissue, of which 15 proteins were shown to be differentially expressed between lean and fat chickens, most of them being involved in lipid metabolism, including adipocyte FABP, apolipoprotein A-I and acyl-coenzyme A dehydrogenase, and oxidative stress, such as HSP 27, which was downregulated, and glutathione-S-transferase- α and S-transferase- β , which were upregulated. Furthermore, members of cytoskeleton family were differentially abundant, underlying a different cytoskeleton rearrangement. These results were partially confirmed by a recent study that investigated the differences in abdominal adipose tissue proteomes between broiler lines divergently selected for abdominal fat content (NEAULF breed) (Wu et al. 2016). Thirteen proteins were found to be differentially expressed between lean and fat lines. In particular, the expression of abdominal adipose tissue Apo A-I, PPIase FKBP4 and cytokeratin otokeratin was found significantly higher in lean birds as compared with fat birds. These proteins are mainly involved in lipid metabolism, amino acid metabolism, signal transduction, energy conversion and antioxidant and cytoskeleton as well as in adipose tissue metabolism.

7.2 Fish Adipose Tissue Proteomics

Lipids are the main source of energy for fish (Weil et al. 2013). Adipose tissue is distributed in fat mainly located around the digestive tract in the abdominal cavity (visceral fat). The percentage of fat in fish may be very high (up to 25% of the body weight). Located on the ventral and dorsal area of the fish is the subcutaneous fat, which is located around the body of the fish, prominently in dorsal and ventral zones. The quality of carcass is dependent on the distribution of adipose tissue between subcutaneous and perivisceral depots, given that adipocytes are also located in muscle myosepta, a connective tissue membrane separating muscle sheets.

Only two proteomics studies have been carried out on fish adipose tissue, both of them on rainbow trout (Oncorhynchus mykiss). Weil et al. (2009) carried out an in vitro investigation on isolated and cultured adipocytes from subcutaneous (dorsal and ventral) and visceral fat depots. The study investigated the differentiated adipocytes from preadipocytes, while other cells that normally compose the adipose tissue depots, such as connective tissues, endothelial cells or blood cells, were excluded. Proteins were separated by means of 2D electrophoresis, and MALDI-TOF identified nine differentially abundant spots depending on differentiation stage. Of these proteins, beta actin and albumin were associated with visceral adipose tissue, whereas annexin, ATP synthase subunit- β , serum deprivation response protein and heart fatty acid-binding protein (H-FABP) were found to be more abundant in subcutaneous AT. The authors found no differences between adipocytes isolated from visceral and subcutaneous adipose tissue, although the amount of protein expressed was different for some of them. Besides proteins related to cell cultivation procedures, such as bovine serum albumin or serum deprivation response protein, the authors identified actin as the more abundant protein in visceral-derived adipocytes. Actin is involved in cytoskeleton structure, and it was found to be overexpressed during adipogenesis of intramuscular bovine adipocytes (Takenouchi et al. 2004) and is involved in cytoskeletal rearrangement as well. ATP synthase, annexin and H-FABP were found to be more abundant in subcutaneous adipocytes, suggesting that this depot is more metabolically active than visceral ones, which is in contrast with results so far demonstrated for mammalian species. The investigation was integrated with histological parameters, such as adipocyte cell diameter determination, presenting the evidence that the diameter of adipocytes derived from visceral AT was higher as compared to those derived from subcutaneous adipose tissue. A study on transcriptomics and proteomics rainbow trout liver focusing on the differences in fat allocation between visceral and muscle AT was carried out after feeding with high- and low-energy diets (Kolditz et al. 2008). The authors sorted the animals in two groups differing for adipose tissue anatomical distribution, namely, those with mainly visceral adipose tissue (lean muscle line) and mainly muscle adipose tissue (fat muscle line). H-FABP was upregulated in the fat muscle cell line when compared with the lean muscle cell line, confirming its importance in fish adipose metabolism.

GAPDH was upregulated as well, whereas apolipoprotein A-1, which is usually overexpressed in adipose tissue, was on the contrary downregulated in the liver.

8 Conclusions: New Insights into the New Knowledge Contributed by System Biology Approach to Better Understanding of Adipose Tissue in Farm Animals

Postgenomic applications in veterinary medicine, including transcriptomics, proteomics and metabolomics, are increasing exponentially. The number of omics studies carried out in farm animals pales if compared to those carried out in humans and a true system biology approach to the impact of omics in livestock are far from being fully implemented. Animal proteomics is further lagging behind those in human biology (Almeida et al. 2015). Before the full development of the "omics" revolution, the knowledge of the physiological bases of adipose tissue development and IMF deposition was obtained through independent analysis of the activity of each single protein. The study of adipose tissue carried out using proteomic approaches, in particular when applied together with a system biology approach, including also transcriptomics and metabolomics, provides an integrated network of the single elements, the knowledge of which provides, in turn, greater information than the sum of individual parts. Transcriptomics and proteomics are an evident leap forward in our understanding of basic biology of adipose tissue, integrating the information necessary to link, for example, adipose tissue and metabolic stress in dairy cows or optimal IMF deposition in cows and pigs. This information might be readily implemented into the field, on the basis that adipose tissue metabolism may be modified by changing fatty acid content through diet. Omics technologies remain quite expensive, proteomics even more. Yet, technology moves rapidly forward and the costs for omics application are constantly dropping. The goal of a \$1000 genome has been almost reached, and it is expected that further drops in omic experiment costs will result in an exponential increase of proteomic studies as well.

Acknowledgements Hereby we acknowledge the support of Dr Helena Arino (he.ar.ba@gmail.com) for the drawing of Fig. 4.

References

- Akter SH, Häussler S, Germeroth D et al (2012) Immunohistochemical characterization of phagocytic immune cell infiltration into different adipose tissue depots of dairy cows during early lactation. J Dairy Sci 95(6):3032–3044
- Almeida AM, Bassols A, Bendixen E et al (2015) Animal board invited review: advances in proteomics for animal and food sciences. Animal 9(1):1–17

- Alvarez-Llamas G, Szalowska E, de Vries MP et al (2007) Characterization of the human visceral adipose tissue secretome. Mol Cell Proteomics 6(4):589–600
- Asano H, Yamada T, Hashimoto O et al (2013) Diet-induced changes in Ucp1 expression in bovine adipose tissues. Gen Comp Endocrinol 184:87–92
- Barbatelli G, Murano I, Madsen L et al (2010) The emergence of cold-induced brown adipocytes in mouse white fat depots is determined predominantly by white to brown adipocyte transdifferentiation. Am J Physiol Endocrinol Metab 298(6):E1244–E1253
- Basse AL, Dixen K, Yadav R et al (2015) Global gene expression profiling of brown to white adipose tissue transformation in sheep reveals novel transcriptional components linked to adipose remodeling. BMC Genomics 16:215
- Bauman DE, Currie WB (1980) Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. J Dairy Sci 63(9):1514–1529
- Bell AW, Bauman DE (1997) Adaptations of glucose metabolism during pregnancy and lactation. J Mammary Gland Biol Neoplasia 2(3):265–278
- Bobe G, Young JW, Beitz DC (2004) Invited review: pathology, etiology, prevention, and treatment of fatty liver in dairy cows. J Dairy Sci 87(10):3105–3124
- Bonnet M, Faulconnier Y, Leroux C et al (2007) Glucose-6-phosphate dehydrogenase and leptin are related to marbling differences among Limousin and Angus or Japanese Black × Angus steers. J Anim Sci 85:2882–2894
- Bonnet M, Cassar-Malek I, Chilliard Y et al (2010) Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. Animal 4:1093–1109
- Bonnet M, Tournayre J, Cassar-Malek I (2016, August) Integrated data mining of transcriptomic and proteomic datasets to predict the secretome of adipose tissue and muscle in ruminants. Mol BioSyst 12(9):2722–2734. https://doi.org/10.1039/c6mb00224b
- Candek-Potokar M, Skrlep M (2012) Factors in pig production that impact the quality of dry-cured ham: a review. Animal 6(2):327–338
- Cánovas A, Quintanilla R, Amills M et al (2010) Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics 11:372
- Cebra CK, Garry FB, Getzy DM et al (1997) Hepatic lipidosis in anorectic, lactating holstein cattle: a retrospective study of serum biochemical abnormalities. J Vet Intern Med 11 (4):231–237
- Ceciliani F, Restelli L, Lecchi C (2014) Proteomics in farm animals models of human diseases. Proteomics Clin Appl 8(9–10):677–688
- Corominas J, Ramayo-Caldas Y, Puig-Oliveras A et al (2013) Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics 14:843
- de Almeida AM, Bendixen E (2012) Pig proteomics: a review of a species in the crossroad between biomedical and food sciences. J Proteome 75(14):4296–4314
- Di Luca A, Elia G, Hamill R, Mullen AM (2013) 2D DIGE proteomic analysis of early post mortem muscle exudate highlights the importance of the stress response for improved waterholding capacity of fresh pork meat. Proteomics 13(9):1528–1544
- Di Luca A, Hamill RM, Mullen AM et al (2016) Comparative proteomic profiling of divergent phenotypes for water holding capacity across the post mortem ageing period in porcine muscle exudate. PLoS One 11(3):e0150605
- Forner F, Kumar C, Luber CA et al (2009) Proteome differences between brown and white fat mitochondria reveal specialized metabolic functions. Cell Metab 10:324–335
- Frayn KN (2002) Adipose tissue as a buffer for daily lipid flux. Diabetologia 45(9):1201–1210
- Gallego M, Mora L, Concepción Aristoy M et al (2016) The use of label-free mass spectrometry for relative quantification of sarcoplasmic proteins during the processing of dry-cured ham. Food Chem 196:437–444
- Geelen MJ, Wensing T (2006) Studies on hepatic lipidosis and coinciding health and fertility problems of high-producing dairy cows using the "Utrecht fatty liver model of dairy cows". A review. Vet Q 28(3):90–104

- Gondret F, Guével B, Com E et al (2012) A comparison of subcutaneous adipose tissue proteomes in juvenile piglets with a contrasted adiposity underscored similarities with human obesity. J Proteome 75(3):949–961
- Griffin TJ, Gygi SP, Ideker T et al (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol Cell Proteomics 1 (4):323–333
- Hammon HM, Stürmer G, Schneider F et al (2009) Performance and metabolic and endocrine changes with emphasis on glucose metabolism in high-yielding dairy cows with high and low fat content in liver after calving. J Dairy Sci 92(4):1554–1566
- Hausman GJ, Basu U, Wei S et al (2014) Preadipocyte and adipose tissue differentiation in meat animals: influence of species and anatomical location. Annu Rev Anim Biosci 2:323–351
- Huang H, Larsen MR, Palmisano G et al (2014) Quantitative phosphoproteomic analysis of porcine muscle within 24 h postmortem. J Proteome 106:125–139
- Imhasly S, Naegeli H, Baumann S et al (2014) Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. BMC Vet Res 10:122
- Kaspric N, Picard B, Reichstadt M et al (2015) ProteINSIDE to easily investigate proteomics data from ruminants: application to mine proteome of adipose and muscle tissues in bovine foetuses. PLoS One 10:e0128086
- Keady SM, Kenny DA, Ohlendieck K et al (2013) Proteomic profiling of bovine M. longissimus lumborum from crossbred Aberdeen Angus and Belgian Blue sired steers varying in genetic merit for carcass weight. J Anim Sci 91(2):654–665
- Kershaw EE, Flier JS (2004) Adipose tissue as an endocrine organ. J Clin Endocrinol Metab 89 (6):2548–2556
- Kim NK, Lee SH, Cho YM et al (2009) Proteome analysis of the m. longissimus dorsi between fattening stages in Hanwoo steer. BMB Rep 42:433–438
- Kolditz CI, Paboeuf G, Borthaire M et al (2008) Changes induced by dietary energy intake and divergent selection for muscle fat content in rainbow trout (Oncorhynchus mykiss), assessed by transcriptome and proteome analysis of the liver. BMC Genomics 9:506
- Lehr S, Hartwig S, Lamers D et al (2012) Identification and validation of novel adipokines released from primary human adipocytes. Mol Cell Proteomics 11(1):M111.010504
- Liu J, Damon M, Guitton N, Guisle I, Ecolan P, Vincent A, Cherel P, Gondret F (2009) Differentially-expressed genes in pig Longissimus muscles with contrasting levels of fat, as identified by combined transcriptomic, reverse transcription PCR, and proteomic analyses. J Agric Food Chem 57(9):3808–3817
- Lizcano F, Vargas D (2016) Biology of beige adipocyte and possible therapy for type 2 diabetes and obesity. Int J Endocrinol 2016:9542061
- Lomiwes D, Farouk MM, Wiklund E, Young OA (2014) Small heat shock proteins and their role in meat tenderness: a review. Meat Sci 96(1):26–40
- Loor JJ, Bionaz M, Drackley JK (2013) Systems physiology in dairy cattle: nutritional genomics and beyond. Annu Rev Anim Biosci 1:365–392
- Louveau I, Perruchot MH, Bonnet M et al (2016) Invited review: pre- and postnatal adipose tissue development in farm animals: from stem cells to adipocyte physiology. Animal 10:1839–1847
- Luo L, Liu M (2016) Adipose tissue in control of metabolism. J Endocrinol 231(3):R77-R99
- Mao YW, Hopkins DL, Zhang YM et al (2016) Beef quality with different intramuscular fat content and proteomic analysis using isobaric tag for relative and absolute quantitation of differentially expressed proteins. Meat Sci 118:96–102
- Mora L, Gallego M, Escudero E et al (2015) Small peptides hydrolysis in dry-cured meats. Int J Food Microbiol 212:9–15
- Mulligan FJ, Doherty ML (2008) Production diseases of the transition cow. Vet J 176(1):3-9
- Nones K, Ledur MC, Ruy DC (2006) Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler \times layer cross. Anim Genet 37(2):95–100

- Oster M, Murani E, Metges CC et al (2011) A high protein diet during pregnancy affects hepatic gene expression of energy sensing pathways along ontogenesis in a porcine model. PLoS One 6 (7):e21691
- Oster M, Murani E, Metges CC et al (2012) A low protein diet during pregnancy provokes a lasting shift of hepatic expression of genes related to cell cycle throughout ontogenesis in a porcine model. BMC Genomics 13:93
- Pioselli B, Paredi G, Mozzarelli A (2011) Proteomic analysis of pork meat in the production of cooked ham. Mol BioSyst 7(7):2252–2260
- Pope M, Budge H, Symonds ME (2014) The developmental transition of ovine adipose tissue through early life. Acta Physiol (Oxford) 210:20–30
- Puig-Oliveras A, Ramayo-Caldas Y, Corominas J et al (2014) Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. PLoS One 9(6):e99720
- Rabe K, Lehrke M, Parhofer KG et al (2008) Adipokines and insulin resistance. Mol Med 14 (11–12):741–751
- Rajesh RV, Heo G, Park M et al (2010) Proteomic analysis of bovine omental, subcutaneous and intramuscular preadipocytes during in vitro adipogenic differentiation. Comp Biochem Physiol Part D Genomics Proteomics 5:234–244
- Ramayo-Caldas Y, Mach N, Esteve-Codina A et al (2012) Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. BMC Genomics 13:547
- Restelli L, Codrea MC, Savoini G et al (2014) LC-MS/MS analysis of visceral and subcutaneous adipose tissue proteomes in young goats with focus on innate immunity and inflammation related proteins. J Proteome 108:295–305
- Romao JM, Jin W, He M et al (2013) Elucidation of molecular mechanisms of physiological variations between bovine subcutaneous and visceral fat depots under different nutritional regimes. PLoS One 8(12):e83211
- Romao JM, He ML, McAllister TA et al (2014) Effect of age on bovine subcutaneous fat proteome: Molecular mechanisms of physiological variations during beef cattle growth. J Anim Sci 92:3316–3327
- Ros-Freixedes R, Gol S, Pena RN et al (2016) Genome-wide association study singles out SCD and LEPR as the two main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs. PLoS One 11(3):e0152496
- Sarr O, Louveau I, Kalbe C et al (2010) Prenatal exposure to maternal low or high protein diets induces modest changes in the adipose tissue proteome of newborn piglets. J Anim Sci 88:1626–1641
- Sarr O, Louveau I, Le Huërou-Luron I et al (2012) Adipose tissue proteomes of intrauterine growth-restricted piglets artificially reared on a high-protein neonatal formula. J Nutr Biochem 23(11):1417–1424
- Sauerwein H, Bendixen E, Restelli L et al (2014) The adipose tissue in farm animals: a proteomic approach. Curr Protein Pept Sci 15(2):146–155
- Sethi JK, Vidal-Puig AJ (2007) Thematic review series: adipocyte biology. Adipose tissue function and plasticity orchestrate nutritional adaptation. J Lipid Res 48(6):1253–1262
- Shen YN, Kim SH, Yoon DH et al (2012) Proteome analysis of bovine longissimus dorsi muscle associated with the marbling score. Asian-Australas J Anim Sci 25:1083–1088
- Shen L, Luo J, Du J et al (2015) Transcriptome analysis of Liangshan pig muscle development at the growth curve inflection point and asymptotic stages using digital gene expression profiling. PLoS One 10(8):e0135978
- Shingfield KJ, Bonnet M, Scollan ND (2013) Recent developments in altering the fatty acid composition of ruminant-derived foods. Animal 7:132–162
- Smith SB, Crouse JD (1984) Relative contributions of acetate, lactate and glucose to lipogenesis in bovine intramuscular and subcutaneous adipose tissue. J Nutr 114:792–800
- Taga H, Chilliard Y, Meunier B et al (2012) Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown? J Cell Physiol 227:1688–1700

- Takenouchi T, Miyashita N, Ozutsumi K et al (2004) Role of caveolin-1 and cytoskeletal proteins, actin and vimentin, in adipogenesis of bovine intramuscular preadipocyte cells. Cell Biol Int 28 (8–9):615–623
- Trayhurn P, Drevon CA, Eckel J (2011) Secreted proteins from adipose tissue and skeletal muscle adipokines, myokines and adipose/muscle cross-talk. Arch Physiol Biochem 117(2):47–56
- Wang D, Wang N, Li N et al (2009) Identification of differentially expressed proteins in adipose tissue of divergently selected broilers. Poult Sci 88(11):2285–2292
- Weil C, Sabin N, Bugeon J et al (2009) Differentially expressed proteins in rainbow trout adipocytes isolated from visceral and subcutaneous tissues. Comp Biochem Physiol Part D Genomics Proteomics 4(3):235–241
- Weil C, Lefevre F, Bugeon J (2013) Characteristics and metabolism of different adipose tissues in fish. J Rev Fish Biol Fisheries 23:157
- Weisber SP, McCann D, Desai M et al (2003) Obesity is associated with macrophage accumulation in adipose tissue. J Clin Invest 112(12):1796–1808
- Wu CY, Wu YY, Liu CD et al (2016) Comparative proteome analysis of abdominal adipose tissues between fat and lean broilers. Proteome Sci 14(1):9
- Xing K, Zhu F, Zhai L et al (2015) Integration of transcriptome and whole genomic resequencing data to identify key genes affecting swine fat deposition. PLoS One 10(4):e0122396
- Xing K, Zhu F, Zhai L et al (2016) Identification of genes for controlling swine adipose deposition by integrating transcriptome, whole-genome resequencing, and quantitative trait loci data. Sci Rep 6:23219
- Xu H, Barnes GT, Yang Q et al (2003) Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J Clin Invest 112(12):1821–1830
- Yang H, XL X, Ma HM, Jiang J (2016) Integrative analysis of transcriptomics and proteomics of skeletal muscles of the Chinese indigenous Shaziling pig compared with the Yorkshire breed. BMC Genet. 17(1):80
- Zachut M (2015) Defining the adipose tissue proteome of dairy cows to reveal biomarkers related to peripartum insulin resistance and metabolic status. J Proteome Res 14(7):2863–2871
- Zhang Y, Proenca R, Maffei M et al (1994) Positional cloning of the mouse obese gene and its human homologue. Nature 372(6505):425–432
- Zhang Q, Lee H-G, Han J-A et al (2010) Differentially expressed proteins during fat accumulation in bovine skeletal muscle. Meat Sci 86:814–820
- Zhang J, Ma J, Long K et al (2016) Dynamic gene expression profiles during postnatal development of porcine subcutaneous adipose. Peer J. 4:e1768
- Zhao YM, Basu U, Dodson MV et al (2010) Proteome differences associated with fat accumulation in bovine subcutaneous adipose tissues. Proteome Sci 8:14

Points clés de l'article 2:

Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gelbased and gel-free proteomic analyses

11 OBJECTIFS

Nous avons identifié des protéines dont l'abondance varie avec la teneur en lipides intramusculaires dans le muscle Longissimus dorsi de bovins de race Rouge des Près, plus tardifs que les races pour lesquelles des données étaient publiées.

Nous avons identifié des protéines dont l'abondance varie avec le pourcentage de tissu adipeux dans la carcasse à partir du protéome du muscle Longissimus dorsi de bovins de race Maine-Anjou.

Nous avons combiné 2 approches protéomiques pour maximiser le nombre de protéines identifiées et se rapprocher des conditions expérimentales déjà publiées.

Enfin nous avons comparé les listes de protéines dont l'abondance varie avec la teneur en lipides intramusculaires dans la race Rouge des Près plus tardive que les races décrites dans l'article 1.



12 DEMARCHE EXPERIMENTALE/ANALYTIQUE

13 CONCLUSIONS

Nous avons identifié 47 protéines différentiellement abondantes entre les 2 groupes de vaches divergentes par leur adiposité musculaire et de carcasse.

Parmi ces protéines, 26 n'avaient jamais été identifiées dans les races plus précoces et elles sont corrélées positivement ou négativement avec les pourcentages de tissus adipeux dans la carcasse ou le muscle (Figure 27).



Figure 27 : Résumé graphique des résultats de l'article. Les protéines qui ont une abondance positivement corrélée aux dépôts adipeux de la carcasse ou du muscle sont en vert tandis que celles qui présentent une corrélation négative sont en rouge. Les protéines qui ont une relation avec le pourcentage de tissu adipeux de la carcasse, du muscle ou les 2 sont indiquées.

Contents lists available at ScienceDirect

Meat Science

journal homepage: www.elsevier.com/locate/meatsci

Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gel-based and gel-free proteomic analyses



MEAT SCIENCE

Jeanne Bazile^a, Brigitte Picard^a, Christophe Chambon^{b,c}, Alberic Valais^d, Muriel Bonnet^{a,*}

^a INRA, Université Clermont Auvergne, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France

^b INRA, UR370, Qualité des Produits Animaux, F-63122 Saint-Genès-Champanelle, France

^c INRA, PFEM cp protéomique, F-63122 Saint-Genès-Champanelle, France

^d S.I.C.A. Rouge des Prés, Domaines des rues, Chenillé Changé, France

ARTICLE INFO

Keywords: Proteome Carcass adiposity Marbling Biomarkers Bovine

ABSTRACT

Adipose tissue (AT) deposits, either intramuscular (i.e., marbling) or at the carcass level, are major economic drivers in the beef industry. To identify biomarkers and unravel mechanisms of AT deposition, we combined gelbased and gel-free methods and we mined the differential proteome of the *longissimus thoracis* from cows differing by two levels of both carcass and muscular adiposities (CMA). We identified 47 proteins with abundances that varied according to CMA. We listed 26 new candidate biomarkers of marbling and confirmed 21 proteins already proposed in the literature. Seven proteins involved in glycolysis or gluconeogenesis were the least abundant, while 14 proteins related to oxidative metabolism, slow-type muscle or retinoic acid metabolism were the most abundant in the high adiposity group. We highlighted eight proteins as differentially abundant and correlated with both CMA, thereby providing the first list of putative biomarkers of carcass adiposity. These proteins would be targeted in future studies aiming to categorize adiposity in cattle.

1. Introduction

Producing meat animals with adequate adipose tissue (AT) masses relative to muscle is an economic challenge for the beef industry. The deposition of AT in the whole body is a key factor for carcass grading and maturation. The deposition of AT within muscle (intramuscular fat, also called marbling and assayed as IMF content) generally positively influences sensory quality traits in beef. This significant economic challenge has motivated research on the growth and metabolism of AT depending on the anatomical site from whole carcass and meat perspectives (Hausman & Canovas, 2017). However, few consensuses on molecular markers of whole-body or muscular adiposity have been proposed (Baik et al., 2017; Ceciliani, Lecchi, Bazile, & Bonnet, 2018). Proteomics offers the opportunity to understand molecular mechanisms involved in AT deposition and to identify the proteins that are differentially abundant related to whole-body AT or IMF deposition and usable as biomarkers. However, to the best of our knowledge, few proteomic analyses have been published to understand or predict the bovine ability to deposit AT at the muscular and whole body levels (Ceciliani et al., 2018). Most of the available proteomic data were obtained from muscles of early maturing breeds (Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Y. Zhang et al., 2011), except one report

of the muscular proteome of cross-breed Belgian blue x Holstein Friesian and Aberdeen Angus x Holstein Friesian cattle (Keady et al., 2013). More recently, differential proteome analyses have identified muscular biomarkers correlated to both carcass quality grade and marbling in beef cattle raised in the Pacific Northwestern United States (Thornton, Chapalamadugu, Eldredge, & Murdoch, 2017) as well as a marbling biomarker in the Bos indicus Nellore breed (De Souza Rodrigues et al., 2017; Poleti et al., 2018). We performed a first mining of the available Bos taurus muscular data, and we established a list of 50 proteins as candidate biomarkers of IMF deposition, of which only 9 have been identified in at least 2 studies (Ceciliani et al., 2018). This slight overlap among putative IMF biomarkers probably results from differences in both breed and proteomic methods. One article identified 27 proteins using a gel-free method (Mao et al., 2016), and five other articles identified 4 to 17 proteins using gel-based methods, either two dimension gel electrophoresis (2-DE) (Kim et al., 2009; Shen et al., 2012; Y. Zhang et al., 2011) or 2D difference gel electrophoresis (2D-DIGE, (Keady et al., 2013)). The gel-based methods developed in the mid-1970s have been a first choice approach in proteome studies ever since, whereas gel-free methods such as the shotgun analysis have been more widely used following the rise of LC-MS or LC-MS/MS (McCormack et al., 1997) in the late 1990s. Due to their methodological

https://doi.org/10.1016/j.meatsci.2019.05.018 Received 7 February 2019; Received in revised form 5 April 2019; Accepted 16 May 2019 Available online 20 May 2019

0309-1740/ © 2019 Elsevier Ltd. All rights reserved.



^{*} Corresponding author. *E-mail address:* muriel.bonnet@inra.fr (M. Bonnet).

peculiarities, gel-based assays offer a focus on the identification of a few specific proteins between experimental conditions and allow analysis of posttranscriptional-modified variants. Shotgun proteomics constitutes a rapid scan of nearly the whole proteome but with less accuracy (Oliveira, Coorssen, & Martins-de-Souza, 2014). Attempts to compare gel-based and gel-free proteomics have repeatedly reported that only a few proteins were common between the two sets of differentially abundant proteins, showing how complementary those two types of assays are (Pieroni et al., 2011).

Whether the putative marbling biomarkers identified by gel-based and gel-free proteomics, mainly in bovines from early-maturing breeds, are reliable in European pure breeds remains to be studied. Moreover, whether the muscular proteome could be mined to search for carcass AT deposition biomarkers was never reported in ruminants, while in humans, cellular and molecular properties of muscle were shown to predict changes in percent body fat (Sun, Ukkola, Rankinen, Joanisse, & Bouchard, 2002). Therefore, the objective of this study was to mine the muscular proteome from Rouge des Pres cows chosen for their high variability in AT deposition to identify proteins that discriminate bovine differences in carcass AT mass and IMF content.

2. Materials and methods

2.1. Animals, handling, slaughtering, and sampling

All animal trials described herein were conducted according to relevant international guidelines (European Union procedures on animal experimentation - Directive 2010/63/EU) for the use of production animals in animal experimentation. Cows (also reared for meat production in France) were reared according to PDO (Protected Designation Origin) recommended production conditions and slaughtered following the commercial practices of the industry (UE Council Regulation (EC) No. 1099/2009) in the same slaughterhouse (ELIVIA, Le Lyon d'Angers 49220, France). Accordingly, no ethics committee or IACUC (Institutional Animal Care and Use Committee) approval was required. The PDO Maine-Anjou official quality mark was involved in the search for marbling indicators to grade meat and carcasses. A total of 20 cows were used (69.73 \pm 27.95 months of age, 423.76 \pm 32.38 kg live weight, Table 1) from the French PDO Maine-Anjou using the "Rouge des Prés" breed, and were representative of the Maine-Anjou cows reared for meat production. These cows were a subset from the trial previously described by Gagaoua, Monteils, Couvreur, & Picard, 2017 chosen for their extreme differences in the percent of both marbling and body fat but with similar rearing and nutritional practices between groups. Groups were composed of cows divergent for both carcass and muscular adiposities and were named high and low adiposity groups (Table 1). Before slaughter, all animals were food-deprived for 24 h and had free access to water. All animals were slaughtered under the same conditions, stunned using a captive-bolt pistol prior to exsanguination and dressed according to standard commercial practices. After slaughter, the hot carcasses were weighed and graded according to the European beef grading system (1249 CE/2008). Muscle samples of the longissimus thoracis (LT) were excised from the fifth rib of the left side of the carcass 24 h after slaughter. Muscle samples were frozen in liquid nitrogen and stored at -80 °C until analysis. The fifth rib was sampled and dissected, and intermuscular fat, meat and bones were separated from each other and weighed. Thereby, the fat to meat ratio of the fifth rib (i.e., the intermuscular fat content) and carcass composition were calculated according to Robelin et al., 1975 (and detailed in Gagaoua, Picard, & Monteils, 2018). The percent of AT in carcasses is reported in Table 1. The IMF content was assayed according to the Soxhlet method using an Accelerated Solvent Extractor 200 from $1 \pm 0,001$ g of meat powder exactly weighted and placed in a 22 mL extraction cell. The extraction was performed with petroleum at a temperature of 125 °C and a pressure of 103 bars. The extract, containing fat content and petroleum ether, was collected and transferred into a previously weighed (\pm 0,001 g) evaporation vial. After 15 min of evaporation, the vial was placed in a drying oven at 105 °C for 17 h and then weighed (\pm 0,001 g) to determine the amount of IMF in the meat sample. The results were expressed as the % of IMF in fresh meat.

2.2. Protein preparation

Protein preparation was performed according to Bouley, Chambon, & Picard, 2004 from the muscle of 20 cows. Briefly, frozen LT muscle tissue (40 mg) was homogenized using a polytron in 1 mL of lysis buffer containing 8.3 M urea, 2 M thiourea, 1% DTT, 2% CHAPS and 2% IPG buffer pH 3–10 and centrifuged at 100006 g for 30 min. The supernatant was harvested, and the protein concentration was determined as described in the RCDC protein assay kit (BioRad, Marnes-la-Coquette, France) with bovine serum albumin as a standard. The supernatant was fractionated and stored at -80 °C pending proteomic analyses. The same protein sample was used for the 2-DE and shotgun assays.

2.3. Two-dimensional gel electrophoresis and mass spectrometry

Two-DE separation was performed according to Bouley et al., 2004 on the protein samples from 20 cows. Briefly, proteins were solubilized in a rehydration solution (8 M urea, 1 M thiourea, 0.28% DTT, 2% CHAPS, 2% IPG buffer pH 4-7, and 0.01% CBB G-250) for isoelectric focusing (IEF). Immobilized pH gradient (IPG) strips covered a pH range of 4–7 and were rehydrated overnight in 340 µl of a protein solution under low-viscosity paraffin oil. IPG strips containing 700 µg of proteins were subjected to IEF (73.5 kVh) at 20.5 °C. Once completed, proteins on the strips were equilibrated for 15 min in a buffer composed of 6 M urea, 1% DTT, 30% glycerol, 50 mM Tris base, 2% SDS and DTT and in the same solution except that DTT was replaced by 5% iodoacetamide for another 15 min. The IPG strips were then transferred onto 12% T, 2.6% C separating polyacrylamide gels, and proteins were separated in the second dimension using a Protean Plus Dodeca cell system (BioRad, Hercules, CA). 2-DE gels were stained using a CBB G-250 procedure and digitized at 300 dpi using a calibrated OD scanner (Expression 10000XL PRO, Epson Corporation, Nagano, Japan). Chemical reagents were from GE Healthcare (Uppsala, Sweden) and Sigma (St.

Table 1

Age of Rouges des Prés cows, carcass and muscular characteristics in the high and low adiposity groups assayed by 2-DE and shotgun proteomics.

| | 2-DE | | | Shotgun | | |
|---|---|--|---|--|--|--------------------------------|
| | High adiposity $(n = 10)$ | Low adiposity $(n = 10)$ | P-value | High adiposity $(n = 5)$ | Low adiposity $(n = 5)$ | P-value |
| Age at slaughter (months) Hot carcass weight (kg) Intramuscular lipid (% IMF) Adipose T. proportion in carcass (%) | $77.05 \pm 32.88 \\ 434.80 \pm 39.23 \\ 5.67 \pm 1.00 \\ 0.33 \pm 0.07$ | $57.42 \pm 12.26 415.30 \pm 22.41 2.71 \pm 0.61 0.23 \pm 0.04$ | $\begin{array}{c} 0.05 \\ 0.10 \\ < 0.01 \\ < 0.01 \end{array}$ | $76.94 \pm 34.31 427.20 \pm 30.01 4.98 \pm 1.64 0.28 \pm 0.05$ | $\begin{array}{rrrrr} 75.10 \ \pm \ 26.86 \\ 399.00 \ \pm \ 22.03 \\ 2.59 \ \pm \ 0.86 \\ 0.20 \ \pm \ 0.03 \end{array}$ | 0.46 0.07 0.01 < 0.01 |

Note: Values are expressed as the means \pm standard deviation.

P-values were calculated by Student's t-test between the groups of high vs. low adiposity within 2-DE or shotgun assays. Adipose T. stands for adipose tissue.

Louis, MO). Protein spot detection and volume quantification were performed using Progenesis SameSpots V.3 software (Nonlinear Dynamics, Newcastle, UK). Protein spots from the 20 animals (groups of high adiposity vs low adiposity) were matched to a reference gel.

Raw volumes of protein spots common to the two groups of adiposity were exported to Excel (Microsoft, Issy-les-Moulineaux, France) and preprocessed to compare proteomes between the adiposity groups (cf Table 1). The data matrix was logged and normalized per adiposity group. Each protein value for one gel (logged volume) was divided by the mean of all protein values for this gel and multiplied by the mean of all protein values for all gels (logged normalized volume). The results were expressed as logged normalized volume, that is, protein abundances.

From these quantifications, 21 spots with an abundance that was or tended to be different between adiposity groups were sampled for identification. Trypsin-digested proteins were analyzed by MALDI-TOF-TOF Autoflex speed spectrometer (Bruker, Marne la Vallée, France) or by nanoESI-IT LTQ velos spectrometer (ThermoFisher Scientific, Courtaboeuf, France) if confirmation of protein identification was necessary. MS or MS/MS spectra were analyzed using ProteinScape V.3 (Bruker) or Proteome Discoverer V1.4.1 (ThermoFisher) and compared to Bos taurus Uniprot database (24,233 sequences) using MASCOT V 2.3 with the parameters adapted to the mass spectrometer used: peptide mass tolerance was set 25 ppm (MALDI-TOF-TOF) or 500 ppm (ESI-IT), fragment mass tolerance was set to 0.2 Da (IT), and a maximum of two missed cleavages was allowed. Variable modifications were methionine oxidation (M) and carbamidomethylation (C) of cysteine. Protein identification was considered valid if at least 3 peptides with a statistically significant Mascot score assigned it (p value > .05 with a False Discovery Rate (FDR) < 5%).

2.4. Shotgun proteomics, LC-MS/MS, protein identification and label-free quantitation

Protein extracted from the LT muscle of 10 cows (chosen among the 20 cows analyzed by 2-DE) was mixed with 1 volume of Laemmli buffer and heated for 5 min at 60°. A fraction of 80 µg of protein was concentrated as two very close bands in the stacking gel (5% polyacrylamide) of an SDS-PAGE (12% acrylamide) gel which was performed using a Mini-Protean II electrophoresis unit (BioRad, Marnes-La-Coquette, France). Gels were run at 80 V during 15 min until the dye front reached the limit zone between the stacking and separating gel (stacking fraction). Gels were stained overnight in Coomassie brilliant blue G-250. Protein bands (the upper band corresponding to high molecular weight myofibrilar proteins, and the lower band corresponding to proteins of lower molecular weights) were cut with a sterile scalpel and reduced with 10 mM DTT for 1 h at 37 °C and subsequently alkylated with 20 mM iodoacetamide for 1 h at RT in the dark. After a wash with 25 mM ammonium bicarbonate, proteins bands were destained in a solution of 25 mM ammonium bicarbonate with acetonitrile $(50/50 \nu/$ v). After dehydration with 100% acetonitrile, gel pieces were dried in a speed vacuum, and samples were preserved at -20 °C until LC-MS/MS analysis.

The hydrolysis of proteins was performed at 37 °C using 800 ng (80 μ L) of sequence grade-modified trypsin/lysC (*Promega, France*) per band. After extraction with 64 μ L of 100% acetonitrile and sonication (15 min), peptides were withdrawn and concentrated in a speed vacuum. The volume was adjusted to 20 μ L with an aqueous solution (99,9% H2O, 0.1% TFA).

Peptide mixture (1 μ L) was injected into the micro HPLC Ultimate 3000, (ThermoFisher Scientific, Courtaboeuf, France) after a preliminary step of desalting and concentration in the pre-column 300 μ m \times 5 mm, (ThermoFisher, Courtaboeuf, France) for 6 min, and a second step of separation in an analytical C18 column 300 μ m, 15 cm, (Acclaimn PepMap, ThermoFisher Scientific, Courtaboeuf, France) with a 6–40% gradient (A: 0.5% FA in water, B: 0.1% FA in 80/20 acetonitrile/water) at 10 μ L/min, 30 °C. The eluate was electrosprayed through the CaptiveSpray ion source into the mass spectrometer QTOF impact II (Bruker, Marne la Vallée, France) operates in CID Data Dependent mode, mass range 150–2000 *m/z*, each MS analysis was succeeded by as many MSMS analysis as possible within 3 s.

At the end of each LC-MS/MS analysis, the raw files (.d) were loaded into the Progenesis QI software (Nonlinear Dynamics, v 4.16, Newcastle, UK), and label-free quantitation was performed using a proprietary workflow (alignment, peak picking, normalization, design set up, quantitation, and protein identification). Label-free quantitation of a protein was achieved if at least two unique peptides of this protein were validated and quantified. The total cumulative abundance of protein was calculated by summing the abundances of all unique peptides allocated to the respective protein. For data searching, Mascot (V.2.5, internally licensed version, www.matrixscience.com) was used with ref_bos_taurus (23,970 sequences; 2018). The following parameters were considered for the searches: peptide mass tolerance was set to 20 ppm, fragment mass tolerance was set to 0.2 Da, and a maximum of two missed cleavages was allowed. Variable modifications were as follows: methionine oxidation (M), carbamidomethylation (C) of cysteine and dehydrated (NO). Protein identification was considered valid if at least two peptides with a statistically significant Mascot score were assigned a false discovery rate (FDR) < 5%.

2.5. Statistical analysis

All of the analyses were performed with R (version 3.3.2). A Shapiro test was performed on protein abundances to check for adherence to normality. Normally or nonnormally distributed data were analyzed by Student's or Kruskal-Wallis tests to identify differences between the high and low adiposity groups (the groups were the same for IMF content and carcass AT mass). Changes in protein abundance between the high and low adiposity groups were considered significant at P < .05, with a tendency at 0.05 < P < .10. Pearson or Spearman correlations were performed respectively, to analyze normally or nonnormally distributed data using the Broom and Hmisc package with the cor.test and rcorr functions. The R script used for these analyses, which was developed for proteomic analyses, is available on GitHub with the DOI https://doi.org/10.5281/zenodo.2539329.

2.6. Gene ontology analysis

To understand the biological functions of the identified proteins, a Gene Ontology (GO) enrichment analysis was performed using ProteINSIDE (http://www.proteinside.org/, (Kaspric, Reichstadt, Picard, Tournayre, & Bonnet, 2015)). Considering the limitation of the GO annotation of genes in the bovine genome, we converted the bovine UniprotIDs to orthologous human EntrezGene IDs using BioMart (http://www.ensembl.org/biomart/), and GO enrichment analysis was performed in the human species.

2.7. Heatmap and dendrogram analyses

The heatmaps were produced using PermuMatrix 1.9.3. (available at http://www.atgc-montpellier.fr/permutmatrix/). The values were transformed into log2 values and then analyzed with the following specifications: the space separator was unticked, the dissimilarity measure was Pearson square distance, the hierarchical classification used inertia according to Ward's method and the lines were centered, meaning the mean of rows was subtracted from each value so that the mean value of each row was 0 (Meunier et al., 2007).

3. Results

3.1. 2-DE proteome

Cows within the low and high adiposity groups differed (P < .01,

J. Bazile, et al.

Table 2

Proteins identified as differentially abundant in Rouge des Prés cows using 2-DE and shotgun proteomics.

| | Uninrot | | | | Correlation | |
|-------------|----------------------|--------------|----------------|---------|-------------------------------|-----------------------------------|
| Assays | Protein accession | Gene Name | Fold Change | P-value | (r) to carcass fat mass | Correlation (r) to IMF content |
| 2-DE -S | P68138 | ACTA1 | 1.21 | <0.01 | 0.73 | 0.76 |
| | Q3ZC09 | ENO3 | 0.62 | < 0.01 | -0.53 | -0.38 |
| | G3MY67 | SDHA | 1.43 | 0.01 | 0.45 | 0.43 |
| | Q2KJH4 | WDR1 | 1,65 | 0.01 | | |
| | P02510 | CRYAB | 1,32 | 0.01 | 0.57 | 0.53 |
| | Q3ZC55 | ACTN2 | 0,53 | 0.02 | | -0.57 |
| | Q13203 | MYBPH | 0,52 | 0.02 | -0.43 | -0.54 |
| | P48644 | ALDH1A1 | 1,27 | 0.02 | | 0.5 |
| | F1MB08 | ENO1 | 1,19 | 0.02 | | 0.43 |
| | G1K1X0 | UQCRC1 | 1,24 | 0.02 | 0.44 | 0.43 |
| | Q3ZC09 | ENO3 | 0,73 | 0.02 | -0.38 | -0.53 |
| | Q5E956 | TPI1 | 0,81 | 0.02 | -0.45 | -0.42 |
| | P15690 | NDUFS1 | 0,54 | 0.03 | -0.45 | -0.40 |
| | F1MX12 | ANKRD2 | 1,24 | 0.03 | | |
| | F1N468 | ADK | 0,85 | 0.04 | | |
| | Q2KJE7 | PSME1 | 0,86 | 0.05 | -0.52 | -0.48 |
| | Q3T0T7 | SAR1B | 0,87 | 0.05 | -0.60 | -0.50 |
| tendancy | P13620 | ATP5H | 0,92 | 0.06 | | -0.42 |
| - | P48644 | ALDH1A1 | 1,20 | 0.07 | 0.45 | |
| | Q8MKH6 | TNNT1 | 1,14 | 0.08 | 0.48 | 0.45 |
| | Q3ZBM1 | ADPRHL1 | 1,21 | 0.08 | | 0.44 |
| Shotaun - S | Q3SYZ8 | PDLIM3 | 1.98 | <0.01 | | 0.75 |
| - | Q2HJ33 | OLA1 | 1.48 | <0.01 | 0.74 | 0.73 |
| | Q3ZC09 | ENO3 | 0.81 | <0.01 | | |
| | Q2KJJ9 | FBP2 | 0.61 | <0.01 | | |
| | Q8MKH6 | TNNT1 | 1.54 | <0.01 | 0.55 | 0.87 |
| | Q3T0P6 | PGK1 | 0.72 | <0.01 | -0.56 | -0.77 |
| | A5D984 | PKM | 0.71 | <0.01 | -0.58 | -0.67 |
| | Q3ZBD7 | GPI | 0.77 | <0.01 | -0.57 | -0.72 |
| | E1BH17 | GSTM1 | 1.34 | 0.01 | 0.66 | 0.81 |
| | P62894 | CYCS | 1.66 | 0.02 | 0.66 | 0.85 |
| | Q3ZBA8 | NDRG2 | 1.89 | 0.03 | 0.58 | 0.72 |
| | Q3SYR3 | APOBEC2 | 1.52 | 0.03 | | 0.95 |
| | Q9BE40 | MYH1 | 0.39 | 0.04 | | -0.57 |
| | P68002 | VDAC2 | 1.28 | 0.04 | | 0.83 |
| | E1BF23 | MYOM2 | 0.61 | 0.05 | | |
| | O97764 | CRYZ | 1.35 | 0.05 | | 0.58 |
| tendancy | F1MME6 | MYOM1 | 0.63 | 0.05 | | |
| | Q01119 | ACTN3 | 0.74 | 0.05 | | |
| | Q9XSC6 | CKM | 1.28 | 0.05 | 0.86 | 0.86 |
| | P48616 | VIM | 1.47 | 0.06 | 0.66 | 0.63 |
| | F1MR86 | FHL1 | 1.32 | 0.07 | | 0.72 |
| | Q3T145 | MDH1 | 1.24 | 0.07 | 0.69 | 0.93 |
| | P02192 | MB | 1.31 | 0.08 | | 0.84 |
| | E1BEL7 | HSPB1 | 1.44 | 0.08 | | |
| | G3MYN5 | TNNI1 | 1.62 | 0.08 | | |
| | G5E507 | HSP90AB1 | 1.17 | 0.08 | | |
| | P02510 | CRYAB | 1.23 | 0.08 | | |
| | A0A140T8A1 | HSPB6 | 1.36 | 0.09 | | |
| | A6H783 | VDAC5P | 0.79 | 0.09 | -0.62 | |
| | F1ME15 | MYL2 | 1.51 | 0.10 | | 0.79 |
| | P01948 | | 1.11 | 0.10 | | |
| | r40044 | ALUHTAT | 1.42 | 0.10 | | |

We reported proteins identified with an abundance significantly (S, p-value < 0.05) different according to the adiposity groups and proteins that tended to be differentially abundant (tendency, 0.05 < p-value < 0.10). Correlation coefficients (r) between the protein abundance and the value of IMF content or carcass fat mass were reported. The significant (p < 0.05) correlation coefficients were written in black over white and highlighted in gray for the coefficients that tended (0.05 < p-value < 0.10) to correlate with the adiposity phenotypes. Fold change values refer to the average abundance of high adipose in comparison to low adipose cows.

Table 1) in carcass adiposity, and the IMF content of the LT muscle was also called marbling. From the muscular proteins extracted from these two groups, a total of 355 two-dimensional spots were detectable, 33 were differentially abundant and 21 proteins were identified with a fold change varying from 0.52 to 1.65 (Supplemental material S1 and Table 2). Of these, the following 17 proteins were significantly differentially abundant (P < .05): myosin binding protein H (MYBPH), α actinin-2 (ACTN2), NADH-ubiquinone oxidoreductase 75 kDa subunit (NDUFS1), 2 protein species of β-enolase (ENO3), triosephosphate isomerase (TPI1), adenosine kinase (ADK), proteasome activator complex subunit 1 (PSME1), GTP-binding protein SAR1b (SAR1B), alphaenolase (ENO1), actin, alpha skeletal muscle (ACTA1), ankvrin repeat domain 2 (ANKRD2), cvtochrome b-c1 complex subunit 1 (UOCRC1), retinal dehydrogenase 1 (ALDH1A1), the α -crystallin B chain (CRYAB), the succinate dehydrogenase [ubiquinone] flavoprotein subunit (SDHA) and the WD repeat-containing protein 1 (WDR1). Four other proteins tended (0.05 < P < .10) to be differentially abundant between the low and high adiposity groups: a protein species of ALDH1A1, troponin T1 (TNNT1), protein ADP-ribosylarginine hydrolase-like protein 1 (ADPRHL1) and ATP synthase subunit d (ATP5H, Table 2).

Compared to the low adiposity group, SAR1B, TPI1, NDUFS1, ENO3, PSME1, ACTN2, MYBPH, ATP5H, ADK were less abundant in the high adiposity group, while TNNT1, ADRPHL1, ENO1, ANKRD2, ALDH1A1, CRYAB, SDHA, WDR1, ACTA1, UQCRC1 were more abundant as visualized on the heatmap (Fig. 1). The most enriched (*p*-value corrected by the Benjamin and Hochberg test ranged from 3.92e-7 to 3.75e-4) GO terms related to biological pathways (BP) that annotated the 10 proteins that were abundant (both at the significant and trend level) in the high adipose group were linked to "muscle contraction" (ANKRD2, ACTA1 CRYAB), "muscle filament sliding" (TNNT1, ACTA1) or to the "oxidation-reduction process" (UQCRC1, SDHA, ALDH1A1) (Fig. 1). The 9 proteins that were more abundant in the low adipose

A - Dendogram

group were annotated by enriched GO terms in the BP category related to glucose metabolism, such as "canonical glycolysis" and "gluconeogenesis" (TPI1, ENO3) or "sarcomere organization" (ACTN2, MYBPH) (Fig. 1).

Among the 21 identified proteins PSME1, ENO3 and SAR1B were significantly negatively correlated with both carcass fat mass and IMF content, while ACTA1 and CRYAB, were positively correlated with both carcass fat mass and IMF content. SDHA was significantly positively correlated with carcass fat mass, while TPI1 and NDFUFS1 were negatively correlated with carcass fat mass. MYBPH and ACTN2 were negatively correlated with IMF content while ALDH1A1 was positively correlated (Table 2).

3.2. Shotgun proteome

Cows within the low and high adiposity groups differed by the carcass adiposity and by the fat content of the LT muscle (P < .05, Table 1). Over the 332 proteins identified by the shotgun proteome, 16 were differentially (P < .05) abundant between the low and high adiposity groups, with a fold change ranging from 0.39 to 1.98. These proteins were PDZ and LIM domain protein 3 (PDLIM3), Obg-like AT-Pase 1 (OLA1), β -enolase (ENO3), fructose-1,6-bisphosphatase isozyme 2 (FBP2), troponin T, slow skeletal muscle (TNNT1), phosphoglycerate kinase 1 (PGK1), pyruvate kinase (PKM), glucose-6-phosphate isomerase (GPI), glutathione S-transferase Mu 1 (GSTM1), cytochrome *c* (CYCS), protein NDRG2 (NDRG2), probable C- > U-editing enzyme APOBEC-2 (APOBEC2), myosin-1 (MYH1), voltage-dependent anion-selective channel protein 2 (VDAC2), myomesin-2 (MYOM2) and zeta-crystallin (CRYZ).

Sixteen other proteins tended (0.05 < P < .10) to be differentially abundant between the low and high adiposity groups: myomesin-1 (MYOM1), alpha-actinin-3 (ACTN3), creatine kinase M-type (CKM),

B - Gene ontology annotation



Fig. 1. :Dendrogram, heatmap representation of the protein abundance (A) and gene ontology annotations (B) for proteins identified as differentially abundant using 2-DE proteomics.

A- One-way hierarchical clustering of standardized (logged, normalized and centered) proteomic data to visually group proteins with an increased (upper part of the heatmap) or decreased (lower part of the heatmap) abundance in the high adiposity group. B- Identification of enriched gene ontology (GO) terms per group. The results obtained with ProteINSIDE and enrichments were expressed as –log10 (p-value) to visually plot them on graphs, which means that a log10 (p-value) of 3, 2, and 1.3 correspond to a p-value of 0.001, 0.01, and 0.05, respectively. In brackets is the total number of proteins within the annotation, and when not too numerous, the gene names are given. Color indicates the GO annotation category: blue indicates cellular components, orange indicates biological pathways and green indicates molecular functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vimentin (VIM), four and a half LIM domains 1 (FHL1), malate dehydrogenase, cytoplasmic (MDH1), myoglobin (MB), heat shock protein beta-1 (HSPB1), troponin I1, slow skeletal type (TNNI1), heat shock protein HSP 90-beta (HSP90AB1), alpha-crystallin B chain (CRYAB), heat shock protein beta-6 (HSPB6), VDAC5P protein (VDAC5P), myosin regulatory light chain 2, ventricular/cardiac muscle isoform (MYL2), tubulin alpha-4A chain (TUBA4A) and Retinal dehydrogenase 1 (ALDH1A1) (Table 1). Their fold changes ranged from 0.63 to 1.62.

Compared to the low adiposity group, MYL2, VDAC2, CRYAB, CRYZ. NDRG2, TNNI1, MDH1, HSPB1, APOBEC2, TUBA4A, CKM, VIM, OLA1, CYCS, MB, PDLIM3, FHL1, HSPB6, TNNT1, GSTM1, ALDH1A1, HSP90AB1, were more abundant in the high adiposity group, while PGK1, FBP2, GPI, MYH1, ENO3, PKM, MYOM2, MYOM1, ACTN3 and VDAC5P were less abundant, as visualized on the heatmap (Fig. 2). The most enriched (p-value corrected by the Benjamin and Hochberg test ranged from 7.35E-14 to 1.82E-3) GO terms in the BP category that annotated the 22 proteins that were abundant (with significance or tendency) in the high adipose group were "muscle filament sliding" (TNNT1, TNNI1, VIM, MYL2), "muscle contraction" (TNNI1, TNNT1, CRYAB), "chaperone-mediated protein folding binding" (HSPB1 HSPB6), "structural constituent of cytoskeleton" (VIM, TUBA4A) and "negative regulation of cell growth" (FHL1, MYL2) (Fig. 2). The 9 proteins that were abundant in the low adipose group were annotated by enriched GO terms in the BP category related to "gluconeogenesis" (GPI, FBP2, PGK1, ENO3), "glycolytic process" (PKM, PGK1, GPI), "structural constituent of muscle" (ACTN3, MYOM1, MYOM2), "skeletal muscle myosin thick filament assembly" (MYOM1, MYOM2) and "muscle alpha-actinin binding" (MYOM1, MYOM2) (Fig. 2).

Among the 32 proteins listed in Table 2, OLA1, CYCS and GSTM1 were significantly positively correlated to both carcass and muscular adiposities. PDLIM3, NDRG2, TNNT1, APOBEC2 and VDAC2 were

A - Dendogram

positively correlated, while PKM, GPI, and PGK1 were negatively correlated with IMF content (Table 2).

Among the 47 proteins identified in the present study, only TNNT1, ENO3, CRYAB and ALDH1A1 were identified as differentially abundant (with significance or tendency) by both 2-DE and shotgun assays.

3.3. Putative biomarkers of marbling

The 2-DE and shotgun assays highlighted 8 and 11 proteins, respectively, as strongly related to IMF, as they were both significantly (P < .05) differentially abundant and correlated with IMF values.

Among the 2-DE identified spots, the protein ALDH1A1 (one of the protein species), ACTA1 and CRYAB were positively correlated, whereas MYBPH, ENO3, PSME1, ACTN2 and SAR1B were negatively correlated with IMF content (Table 3). Among the 11 proteins identified by the shotgun assay that we proposed to be strongly related to IMF deposition, APOBEC2, OLA1, CYCS, GSTM1, NDRG2, TNNT1, PDLIM3 and VDAC2 were positively correlated, while PKM, GPI and PGK1 were negatively correlated with IMF values. We proposed these 19 proteins as putative reliable biomarkers of IMF content, with APOBEC2 abundance showing the strongest correlation with IMF values (Fig. 3). Sixteen proteins were listed (Table 3) as additional putative biomarkers because they were either differentially abundant according to or correlated with IMF values.

3.4. Putative biomarkers of carcass adiposity

The 2-DE and the shotgun assays highlighted 8 and 3 proteins, respectively, as strongly related to carcass adiposity content as they were both significantly (P < .05) differentially abundant and correlated to fat carcass values.



B - Gene ontology annotation

Fig. 2. Dendrogram, heatmap representation of the protein abundance (A) and gene ontology annotations (B) for proteins identified as differentially abundant using shotgun proteomics.

One-way hierarchical clustering of standardized (logged, normalized and centered) proteomic data to visually group proteins with an increased (upper part of the heatmap) or decreased (lower part of the heatmap) abundance in the high adiposity group. B-Identification of enriched gene ontology terms per group. The results obtained with ProteINSIDE and enrichments were expressed as –log10 (p-value) to visually plot them on graphs, which means that a log10 (p-value) of 3, 2, and 1.3 correspond to a p-value of 0.001, 0.01, and 0.05, respectively. In brackets is the total number of proteins within the annotation, and when not too numerous, the gene names are given. Color indicates the GO annotation category: blue indicates cellular components, orange indicates biological pathways and green indicates molecular functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

 Table 3

 List of putative muscular biomarkers of IMF or carcass fat mass.

| Proteins | Carcass fat mass | IMF | Methods |
|----------|------------------------|-----|---------|
| ACTA1 | S | S | 2-DE |
| CRYAB | S | S | 2-DE |
| CYCS | S | S | Shotgun |
| ENO3 | S | S | 2-DE |
| GSTM1 | S | S | Shotgun |
| OLA1 | S | S | Shotgun |
| PSME1 | S | S | 2-DE |
| SAR1B | S | S | 2-DE |
| NDUFS1 | S | Т | 2-DE |
| SDHA | S | Т | 2-DE |
| TPI1 | S | Т | 2-DE |
| ALDH1A1 | Т | S | 2-DE |
| GPI | Т | S | Shotgun |
| MYBPH | Т | S | 2-DE |
| NDRG2 | Т | S | Shotgun |
| PGK1 | Т | S | Shotgun |
| PKM | Т | S | Shotgun |
| TNNT1 | Т | S | Shotgun |
| CKM | Т | Т | Shotgun |
| MDH1 | Т | Т | Shotgun |
| TNNT1 | Т | Т | 2-DE |
| UQCRC1 | Т | Т | 2-DE |
| VIM | Т | Т | Shotgun |
| VDAC5P | Т | | Shotgun |
| ACTN2 | | S | 2-DE |
| APOBEC2 | | S | Shotgun |
| PDLIM3 | | S | Shotgun |
| VDAC2 | | S | Shotgun |
| ADPRHL1 | | Т | 2-DE |
| ATP5H | | Т | 2-DE |
| CRYZ | | Т | Shotgun |
| ENO1 | | Т | 2-DE |
| FHL1 | | Т | Shotgun |
| MB | | Т | Shotgun |
| MYH1 | | Т | Shotgun |
| MYL2 | | т | Shotaun |

Light grey indicates a positive correlation, whereas dark grey indicates a negative correlation with muscular and/or carcass adiposities. S represents putative reliable biomarkers defined as both significantly differentially abundant according to and correlated with adiposity values. T represents putative biomarkers defined as either significantly differential or significantly correlated with adiposity values.

From the 2-DE assay, ACTA1, CRYAB, SDHA were positively correlated, whereas TPI1, PSME1, ENO3, NDUFS1 and SAR1B were negatively correlated with carcass adiposity values. Among the shotgun-



Fig. 3. Correlation between the abundance of the APOBEC2 protein and the intramuscular fat content (IMF content).

identified proteins, 3 were strongly positively correlated with fat carcass deposition and IMF deposition, as shown before: OLA1, CYCS and GSTM1. We proposed these 11 proteins as putative reliable biomarkers of carcass adiposity content.

3.5. Putative biomarkers of both marbling and carcass adiposity

The 2-DE and the shotgun assays highlighted 5 and 3 proteins, respectively, as strongly related to both marbling and carcass adiposity content as they were both significantly (P < .05) differentially abundant and correlated to fat carcass and intramuscular values. Of these ACTA1and CRYAB (identified by 2-DE) as well as CYCS, GSTM1 and OLA1 (identified by shotgun) were positively correlated with adiposity values, while PSME1, ENO3 and SAR1B (identified by 2-DE) were negatively related to adiposity values.

4. Discussion

We used two proteomic methods to screen for differences in the muscular proteome of cows differing in both carcass adiposity and marbling. We listed 47 proteins that were, or tended to be, differentially abundant between the high and low adiposity groups. Of these, only four proteins (TNNT1, ENO3, CRYAB and ALDH1A1) were identified by both methods as differentially abundant, even if we have to note that similar pathways were identified by the two methods. This low overlap in the protein identified may partly result from the identification of exclusively acidic proteins by 2-DE, while both acidic and basic proteins may be identified by shotgun. The number of proteins that we proposed to be related to carcass and muscular fat deposition is higher than the number of proteins usually reported by differential proteomic studies in bovine muscle (Keady et al., 2013; Mao et al., 2016; Shen et al., 2012; Y. Zhang et al., 2011), probably because of the use of two different proteomic methods, in our case 2-DE and label-free proteomics. Pieroni et al., 2011 and Kubota, Kosaka, & Ichikawa, 2005 have already reported the added value of using two proteomic methods since they target proteins differently depending on their molecular weight, pH, solubilization, and concentration range. In these papers, as in the present study, there were very few common proteins between the two methods, and some of the identified proteins, especially protein species

(similar proteins originating from a single gene and produced by alternative splicing, variable promoter usage or other posttranscriptional modifications (Schlüter, Apweiler, Holzhütter, & Jungblut, 2009)), were exclusively identified by 2-DE assay. By 2-DE, we identified two isoforms of *ALDH1A1* and *ENO3*, among the 5 and 13 isoforms known in primates, respectively, but these are not yet identified in bovines.

4.1. Putative biomarkers of marbling

Among the 47 proteins differentially abundant according to marbling, 21 proteins (ACTA1, ACTN3, ALDH1A1, APOBEC2, ATP5H, CRYAB, ENO1, ENO3, GPI, HSPB1, MB, MDH1, MYBPH, MYH1, MYL2, PGK1, PKM, SDHA, TNNT1, TPI1, VIM) were previously revealed by proteome screening of the longissimus lumborum and/or longissimus thoracis divergent in marbling as the result of age (Kim et al., 2009; Q. Zhang et al., 2010), breed (De Souza Rodrigues et al., 2017; Keady et al., 2013), or marbling score (Thornton et al., 2017) at 24 (Poleti et al., 2018; Shen et al., 2012) (the latter for the Nellore breed) or 30 (Mao et al., 2016) months of age. Of these, HSPB1(Keady et al., 2013; Mao et al., 2016; Thornton et al., 2017; Q. Zhang et al., 2010), TPI1(Kim et al., 2009; Mao et al., 2016; Shen et al., 2012; Thornton et al., 2017), PGK1(Mao et al., 2016; Poleti et al., 2018; Shen et al., 2012; Thornton et al., 2017), SDHA (Kim et al., 2009; Poleti et al., 2018; Thornton et al., 2017), MYH1(Mao et al., 2016; Thornton et al., 2017), ACTA1(De Souza Rodrigues et al., 2017; Thornton et al., 2017), ENO1(De Souza Rodrigues et al., 2017; Thornton et al., 2017), ENO3(Poleti et al., 2018; Thornton et al., 2017) and TNNT1 (Keady et al., 2013; Thornton et al., 2017) were among the potential robust biomarkers of marbling, as they are found in Rouge des Prés cows and in at least two other breeds, with either positive or negative relation to marbling, according to previous studies. The present study also confirms the previously reported lower abundance of ACTN3 (Keady et al., 2013) and MYBPH (Poleti et al., 2018) and the higher abundance of VIM (Mao et al., 2016) and ACTA1 in heifers (Thornton et al., 2017) with highly compared to lightly marbled muscles. Conversely, we reported an opposite relationship between IMF content and APOBEC2, CRYAB, GPI, MB, MYL2, ALDH1A1, ATP5H, MDH1 abundances in Rouge des Prés cows (present study) in comparison with results from the longissimus lumborum or thoracis of steers of different breeds, such as Korean, Xiangxi yellow \times Angus, Aberdeen Angus \times Holstein Friesian, and Nellore (De Souza Rodrigues et al., 2017; Keady et al., 2013; Mao et al., 2016; Poleti et al., 2018; Thornton et al., 2017; Y. Zhang et al., 2011). Among the 47 proteins listed in the present study, we proposed 19 candidate biomarkers of marbling that we considered reliable because they were both significantly differentially abundant between marbling groups and correlated with IMF content in either 2-DE or shotgun assays (Table 3). Knowledge of the biological functions of these proteins also agrees with their link with marbling. For example, ALDH1A1 is a retinal dehydrogenase that irreversibly oxidizes retinaldehyde to retinoic acid (Takeda et al., 2016). Retinoic acid stimulates lipid deposition, whereas retinaldehyde has an inhibitory effect in human adipose-derived stem cells (Takeda et al., 2016). In progenitor cells present in calf muscle, retinoic acid was proposed to stimulate adipogenesis and lipogenesis through ZFP423 and PPARy signaling (Harris et al., 2018). In mature adipocytes, retinoic acid promotes lipid oxidation through the peroxisome proliferator-activated receptor β/δ (Wang, Yang, Harris, Nelson, & Busboom, 2017). In line with this, we may also speculate that an ALDH1A1-mediated increase in retinoic acid content has contributed to the increase in the abundance of the small heat shock protein CRYAB, which is a retinoic acid-responsive gene (Takeda et al., 2016). The higher abundance of CRYAB may sustain the higher oxidative metabolism of highly marbled muscle, which agrees with the higher abundance of both ALDH1A1 and CRYAB in oxidative compared to glycolytic muscles (Picard et al., 2018) and with the role of CRYAB in mitochondrial homeostasis and mitoprotection (Diokmetzidou et al., 2016). Moreover, CRYAB had been shown to interact with some oxidative enzymes, such as VDAC (Diokmetzidou et al., 2016), which was also identified as more abundant in the high adipose group.

The reliability of the marbling biomarkers identified in this study is also supported by the identification of proteins such as NDGR2, APOBEC2 or VDAC that would signal an increased oxidative capacity. NDGR2 protein is highly expressed in muscle under the control of the peroxisome proliferator-activated receptor-gamma coactivator-1a (PGC-1a), a key transcriptional factor of mitochondrial biogenesis and oxidative metabolism (Foletta et al., 2013). We determined an increased abundance of NDGR2 in highly versus lightly marbled muscle that may have inhibited glycolysis if we assume that this regulation. first described in colorectal (Xu et al., 2015) and renal (Shi et al., 2017) cells, occurs in muscle. APOBEC2, a muscle-specific protein that is part of the activation-induced deaminase/apolipoprotein B mRNA editing enzyme catalytic polypeptide cytidine deaminase family, is preferentially found in slow-twitch muscle and appears to be important in fiber type determination. APOBEC2 was previously reported as more abundant in muscle of high marbled bovine (Thornton et al., 2017), however its potential role in marbling development was never discussed. From evidences reported in mice, we proposed that APOBEC2 is either involved or marks, an higher proportion of oxidative fibers repeatedly shown to contain more lipids than glycolytic ones. Indeed, a deficiency of APOBEC2 in mice leads to a tendency for a greater ratio of slow to fast twitch muscle, a reduction in body mass, partially explained by a reduction of the mass of skeletal muscle (Sato et al., 2010), and a mitochondrial defect that increases muscle mitophagy (Sato et al., 2018). From these results, APOBEC2 was proposed as a key molecular driver of mitochondrial homeostasis to maintain normal skeletal muscle function. Whether the strong correlation that we observed between APOBEC2 and IMF content indicates a higher amount of slow-oxidative type muscle or a higher muscular uptake of lipoprotein-associated lipids remains to be studied. VDAC2 forms a channel through the mitochondrial outer membrane that allows the exit of ATP and the entry of calcium (Shimizu et al., 2015; Subedi et al., 2011), a positive regulator of the Krebs cycle enzyme isocitrate dehydrogenase (Denton & McCormack, 1985; Traaseth, Elfering, Solien, Haynes, & Giulivi, 2004) activity that was reported to increase concomitantly to marbling (Jurie et al., 2007). As in the present study, VDAC2 has been shown to be differentially abundant during marbling deposition in chickens, concomitantly to several other proteins involved in complexes 1, 2 and 5 of the respiratory chain (Liu et al., 2017).

In summary, over the 47 muscular proteins differentially abundant according to the adiposity groups, 21 of them were identified both in Rouge des Prés and in at least one other bovine breed. Additionally, we proposed 19 potential biomarkers of marbling in Rouge des Prés cows. Their reliability, which remains to be confirmed in a larger bovine population, is supported by knowledge of the pathways sustained by these proteins, as exemplified by the retinoic acid-related pathways or those driving a shift from glycolytic to oxidative metabolism in the most marbled muscle.

4.2. Putative biomarkers of carcass adiposity

Several striking pieces of evidence for developmental and functional links between muscle and adipose tissue were repeatedly reviewed (Bonnet, Cassar-Malek, Chilliard, & Picard, 2010; Hausman, Basu, Wei, Hausman, & Dodson, 2014; Louveau, Perruchot, Bonnet, & Gondret, 2016) and indicate that muscle metabolism may impact adipose tissue deposition. A well-known bovine model is the myostatin mutation that in double-muscled Belgian blue cattle increases muscular development while it decreases carcass and muscular adipose tissue development (Bonnet et al., 2010). From this evidence, we hypothesized that the abundance of some muscular proteins may help to predict carcass adiposity due to their involvement in nutrient utilization and partitioning between tissues or as regulatory proteins of the balance between
muscle and adipose cells. In agreement with this hypothesis, in humans, muscular signatures such as the proportion of type I muscle fibers or the ratio of the activities of a glycolytic to an oxidative enzyme were related to variation in body weight or body composition, including fat mass (Sun et al., 2002). In bovines, Thornton et al., 2017 reported 45 unique muscular proteins as differentially abundant between high- and low-quality graded beef carcasses; nevertheless these were only divergent based on the marbling score.

Among the 47 proteins differentially abundant according to the adiposity groups, we proposed 11 reliable biomarkers of carcass adiposity (ACTA1, CRYAB, CYCS, GSTM1, OLA1, PSME1, SAR1B, SDHA, ENO3, TPI1, NDUFS1) because they were both significantly differentially abundant between adipose groups and correlated with carcass AT mass content in either 2-DE or shotgun assays (Table 3). Of these, 8 were also among the 19 putative biomarkers of marbling and would mainly indicate a higher oxidative than glycolytic metabolism, at least in muscle. Moreover, of the 3 remaining proteins that we proposed as putative specific biomarkers of carcass adiposity, namely, TPI1, NDUFS1 and SDHA, are also involved in oxidative or glycolytic pathways. In agreement with these results from studies in bovine, in humans, the ratio of a glycolytic enzyme to a tricarboxylic acid cycle enzyme, specifically, phosphofructokinase/ oxoglutarate dehydrogenase, was reported to correlate with the change of the whole-body fat mass during overfeeding (Sun et al., 2002). If we assume that the decreased abundance of glycolytic enzymes in muscle reflects a lesser muscular utilization of glucose, this may have resulted in a higher glucose availability at the whole body level that would promote carcass adiposity, since glucose was shown to be a strong stimulator of lipogenesis in ruminants (Prior & Scott, 1980). Moreover, when glucose availability and the rate of lipogenesis increase, Smith et al., 2018, have recently shown that glucose is replaced by acetate as a significant source of carbon for fatty acid biosynthesis both in muscular and subcutaneous/carcass adipose tissues.

More generally, the muscular molecular signature of high compared to low adipose cows showed a decrease in abundance of proteins related to glucose metabolism and fast twitching, while the abundance of some proteins related to oxidative metabolism was increased. Indeed, over the 47 listed proteins discriminating the two adiposity groups, 7 proteins (PGK1, ENO3, ENO1, GPI, TPI1, PKM, FBP2) were involved in glycolysis, gluconeogenesis or glycogenesis, and their abundance was lower in the high adiposity group. At the opposite, MDH1, SDHA, NDUFS1, UOCRC1 and VDAC2 invoved in oxidative metabolisms were more abundant in the high than in the low marbled Rouge des Prés cows. Contractile proteins such as MYH1 (coding for MyHC-2 \times isoform), related to fast glycolytic fiber, and MyBPH were less abundant in muscle from the high adiposity group, while in this group, we observed a higher abundance of slow-type proteins such as TNNT1 and MYL2. In summary, we proposed 11 muscular putative biomarkers of carcass adiposity that suggested a higher amount of oxidative rather than glycolytic metabolism in muscle which may have increased the glucose availability for non-muscular tissue, an activator of whole body lipogenesis in bovines. To the best of our knowledge, this is the first study to propose muscular proteins as indicators of carcass adiposity in bovines.

5. Conclusion

To conclude, by combining two proteomic methods, we identified 47 proteins that were differentially abundant among two groups of cows divergent in both carcass and muscular adiposities. Of these, 26 were never disclosed by differential proteomic screening according to marbling score. Proteomic changes that occurred in the fatter cows indicated increased oxidative metabolism, decreased glycolytic metabolism, increased retinoic acid metabolism, and probable regulation of cortical actin depolymerization/stabilization with both sarcomeric and signaling impacts. It is noteworthy that the putative contribution of the identified proteins to pathways of muscular or carcass adiposity often relies on the understanding of human and murine species. Regardless, the results obtained in this study expand our understanding of muscular proteome dynamics in light of adiposity, which might help with the development of protein quantitation methods (such targeted proteomics or antibody-based proteomics) to assay or predict carcass or muscular adiposity in beef cattle in the future.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meatsci.2019.05.018.

Contributions

JB and MB drafted the manuscript. MB, BP and AV participated in the experimental design. CC provided the mass spectrometry results. All authors collaborated with the interpretation and discussion of the results. All authors have given approval to the final versions of the manuscript.

Funding sources

This work was supported by the regional council of Pays de Loire (Quafima). J. Bazile was granted by the regional council of Auvergne-Rhône-Alpes and the Inra-Phase Division.

Conflict of interests

The authors declared no conflicts of interest.

Acknowledgements

The authors thank the SICA Rouge des Prés, especially G. Aminot, for animal handling. We thank S. Couvreur and G. Le Bec (Ecole Supérieure d'Agriculture (ESA) Angers) for muscle sampling and IMF content assays, C. Barboiron (INRA) for technical assistance and help on proteomics analyses, for and I. Molnar for advice and writing part of the statistical script used to analyze these results, and B. Meunier for advice in the use of Permutmatrix.

References

- Baik, M., Kang, H. J., Park, S. J., Na, S. W., Piao, M., Kim, S. Y., ... Moon, Y. S. (2017). Triennial growth and development symposium: Molecular mechanisms related to bovine intramuscular fat deposition in the longissimus muscle. *Journal of Animal Science*, 95(5), 2284–2303. https://doi.org/10.2527/jas2016.1160.
- Bonnet, M., Cassar-Malek, I., Chilliard, Y., & Picard, B. (2010). Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal*, 4(7), 1093–1109. https://doi.org/10.1017/S1751731110000601.
- Bouley, J., Chambon, C., & Picard, B. (2004). Mapping of bovine skeletal muscle proteins using two-dimensional gel electrophoresis and mass spectrometry. *Proteomics*, 4(6), 1811–1824. https://doi.org/10.1002/pmic.200300688.
- Ceciliani, F., Lecchi, C., Bazile, J., & Bonnet, M. (2018). Proteomics research in the adipose tissue. Proteomics in domestic animals: From farm to systems biology (pp. 233–254). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69682-9 12.
- De Souza Rodrigues, R. T., Chizzotti, M. L., Vital, C. E., Baracat-Pereira, M. C., Barros, E., Busato, K. C., ... Da Silva Martins, T. (2017). Differences in beef quality between Angus (Bos taurus taurus) and Nellore (Bos taurus indicus) cattle through a proteomic and phosphoproteomic approach. *PLoS ONE*, 12(1), 1–21. https://doi.org/10.1371/ journal.pone.0170294.
- Denton, R. M., & McCormack, J. G. (1985). Ca2+ transport by mammalian mitochondria and its role in hormone action. *The American Journal of Physiology*, 249(6 Pt 1), E543–E554. https://doi.org/10.1080/02699930143000248.
- Diokmetzidou, A., Soumaka, E., Kloukina, I., Tsikitis, M., Makridakis, M., Varela, A., ... Capetanaki, Y. (2016). Desmin and αB-crystallin interplay in the maintenance of mitochondrial homeostasis and cardiomyocyte survival. *Journal of Cell Science*, 129(20), 3705–3720. https://doi.org/10.1242/jcs.192203.
- Foletta, V. C., Brown, E. L., Cho, Y., Snow, R. J., Kralli, A., & Russell, A. P. (2013). Ndrg2 is a PGC-1α/ERRα target gene that controls protein synthesis and expression of contractile-type genes in C2C12 myotubes. *Biochimica et Biophysica Acta, Molecular Cell Research*, 1833(12), 3112–3123. https://doi.org/10.1016/j.bbamcr.2013.08. 011.
- Gagaoua, M., Monteils, V., Couvreur, S., & Picard, B. (2017). Identification of biomarkers associated with the rearing practices, carcass characteristics, and beef quality: An

J. Bazile, et al.

- Gagaoua, M., Picard, B., & Monteils, V. (2018). Associations among animal, carcass, muscle characteristics, and fresh meat color traits in Charolais cattle. *Meat Science*, 140(March), 145–156. https://doi.org/10.1016/j.meatsci.2018.03.004.
- Harris, C. L., Wang, B., Deavila, J. M., Busboom, J. R., Maquivar, M., Parish, S. M., ... Du, M. (2018). Vitamin a administration at birth promotes calf growth and intramuscular fat development in Angus beef cattle. *Journal of Animal Science and Biotechnology*, 9(1), 1–9. https://doi.org/10.1186/s40104-018-0268-7.
- Hausman, G. J., Basu, U., Wei, S., Hausman, D. B., & Dodson, M. V. (2014). Preadipocyte and adipose tissue differentiation in meat animals: Influence of species and anatomical location. *Annual Review of Animal Biosciences*, 2(1), 323–351. https://doi.org/ 10.1146/annurev-animal-022513-114211.
- Hausman, G. J., & Canovas, A. (2017). 2016 triennial growth and development symposium: New perspectives and insight on intramuscular adipose tissue. *Journal of Animal Science*, 95(5), 2241–2243. https://doi.org/10.2527/jas2017.1480.
- Jurie, C., Picard, B., Hocquette, J. F., Dransfield, E., Micol, D., & Listrat, A. (2007). Muscle and meat quality characteristics of Holstein and Salers cull cows. *Meat Science*, 77(4), 459–466. https://doi.org/10.1016/j.meatsci.2007.04.014.
- Kaspric, N., Reichstadt, M., Picard, B., Tournayre, J., & Bonnet, M. (2015). Protein function easily investigated by genomics data mining using the ProteINSIDE online tool. *Genomics and Computational Biology*, 1(1), 16. https://doi.org/10.18547/gcb. 2015.voll.iss1.e16.
- Keady, S. M., Kenny, D. A., Ohlendieck, K., Doyle, S., Keane, M. G., & Waters, S. M. (2013). Proteomic profiling of bovine M. longissimus lumborum from crossbred Aberdeen Angus and Belgian blue sired steers varying in genetic merit for carcass weight. *Journal of Animal Science*, 91(2), 654–665. https://doi.org/10.2527/jas.2012-5850.
- Kim, N. K., Lee, S. H., Cho, Y. M., Son, E. S., Kim, K. Y., Lee, C. S., ... Park, E. W. (2009). Proteome analysis of the m. longissimus dorsi between fattening stages in Hanwoo steer. *BMB Reports*, 42(7), 433–438. https://doi.org/10.5483/BMBRep.2009.42.7. 433.
- Kubota, K., Kosaka, T., & Ichikawa, K. (2005). Combination of two-dimensional electrophoresis and shotgun peptide sequencing in comparative proteomics. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, 815(1–2), 3–9. https://doi.org/10.1016/j.jchromb.2004.10.030.
- Liu, R., Wang, H., Liu, J., Wang, J., Zheng, M., Tan, X., ... Wen, J. (2017). Uncovering the embryonic development-related proteome and metabolome signatures in breast muscle and intramuscular fat of fast-and slow-growing chickens. *BMC Genomics*, 18(1), 1–15. https://doi.org/10.1186/s12864-017-4150-3.
- Louveau, I., Perruchot, M. H., Bonnet, M., & Gondret, F. (2016). Invited review: Pre- and postnatal adipose tissue development in farm animals: From stem cells to adipocyte physiology. *Animal*, 10(11), 1839–1847. https://doi.org/10.1017/ \$1751731116000872.
- Mao, Y., Hopkins, D. L., Zhang, Y., Li, P., Zhu, L., Dong, P., ... Luo, X. (2016). Beef quality with different intramuscular fat content and proteomic analysis using isobaric tag for relative and absolute quantitation of differentially expressed proteins. *Meat Science*, 118, 96–102. https://doi.org/10.1016/j.meatsci.2016.03.028.
- McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., & Yates, J. R. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-Femtomole level. *Analytical Chemistry*, 69(4), 767–776. https://doi.org/10.1021/ac960799q.
- Meunier, B., Dumas, E., Piec, I., Béchet, D., Hébraud, M., & Hocquette, J.-F. (2007). Assessment of hierarchical clustering methodologies for proteomic data mining. *Journal of Proteome Research*, 6(1), 358–366. https://doi.org/10.1021/pr060343h.
- Oliveira, B. M., Coorssen, J. R., & Martins-de-Souza, D. (2014). 2DE: The Phoenix of proteomics. Journal of Proteomics, 104, 140–150. https://doi.org/10.1016/j.jprot. 2014.03.035.
- Picard, B., Gagaoua, M., Al-Jammas, M., De Koning, L., Valais, A., & Bonnet, M. (2018). Beef tenderness and intramuscular fat proteomic biomarkers: Muscle type effect. *PeerJ*, 6, e4891. https://doi.org/10.7717/peerj.4891.
- Pieroni, L., Finamore, F., Ronci, M., Mattoscio, D., Marzano, V., Mortera, S. L., ... Urbani, A. (2011). Proteomics investigation of human platelets in healthy donors and cystic fibrosis patients by shotgun nUPLC-MS ^E and 2DE: A comparative study. *Molecular BioSystems*, 7(3), 630–639. https://doi.org/10.1039/C0MB00135J.
- Poleti, M. D., Regitano, L. C. A., Souza, G. H. M. F., Cesar, A. S. M., Simas, R. C., Silva-Vignato, B., & Coutinho, L. L. (2018). Longissimus dorsi muscle label-free quantitative proteomic reveals biological mechanisms associated with intramuscular fat

deposition. Journal of Proteomics, 179(November 2017), 30-41. https://doi.org/10.1016/j.jprot.2018.02.028.

- Prior, R. L., & Scott, R. A. (1980). Effects of intravenous infusions of glucose, lactate, propionate or acetate on the induction of Lipogenesis in bovine adipose tissue. *The Journal of Nutrition*, 110(10), 2011–2019. https://doi.org/10.1093/jn/110.10.2011.
- Robelin, J., Geay, Y., Jailler, R., Cuylle, G., Carcasses, D. E. S., Bovins, D. E. J., & Un, C. D. (1975). Estimation de la composition des carcasses de jeunes bovins a partir de la composition d'un morceau monocostal ´au niveau de la 11e c ote. PR ELEV.
- Sato, Y., Ohtsubo, H., Nihei, N., Kaneko, T., Sato, Y., Adachi, S. I., ... Yoshizawa, F. (2018). Apobec2 deficiency causes mitochondrial defects and mitophagy in skeletal muscle. *FASEB Journal*, 32(3), 1428–1439. https://doi.org/10.1096/fj.201700493R.
- Sato, Y., Probst, H. C., Tatsumi, R., Ikeuchi, Y., Neuberger, M. S., & Rada, C. (2010). Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. *Journal of Biological Chemistry*, 285(10), 7111–7118. https://doi.org/ 10.1074/jbc.M109.052977.
- Schlüter, H., Apweiler, R., Holzhütter, H. G., & Jungblut, P. R. (2009). Finding one's way in proteomics: A protein species nomenclature. *Chemistry Central Journal*, 3(1), 1–10. https://doi.org/10.1186/1752-153X-3-11.
- Shen, Y. N., Kim, S. H., Yoon, D. H., Lee, H. G., Kang, H. S., & Seo, K. S. (2012). Proteome analysis of bovine Longissimus dorsi muscle associated with the marbling score. *Asian-Australasian Journal of Animal Sciences*, 25(8), 1083–1088. https://doi.org/10. 5713/ajas.2012.12127.
- Shi, W., Xu, X., Yan, F., Wang, B., Zhao, H., Chan, A., & Yuan, J. (2017). N-Myc downstream-regulated gene 2 restrains glycolysis and glutaminolysis in clear cell renal cell carcinoma. Oncology Letters, 14(6), 6881–6887. https://doi.org/10.3892/ol.2017. 7024.
- Shimizu, H., Schredelseker, J., Huang, J., Lu, K., Naghdi, S., Lu, F., ... Chen, J. N. (2015). Mitochondrial ca2+uptake by the voltage-dependent anion channel 2 regulates cardiac rhythmicity. *ELife*, 2015(4), 1–20. https://doi.org/10.7554/eLife.04801.
- Smith, S. B., Blackmon, T. L., Sawyer, J. E., Miller, R. K., Baber, J. R., Morrill, J. C., & Wickersham, T. A. (2018). Glucose and acetate metabolism in bovine intramuscular and subcutaneous adipose tissues from steers infused with glucose, propionate, or acetate. *Journal of Animal Science*, 96(3), 921–929. https://doi.org/10.1093/jas/ sky017.
- Subedi, K. P., Kim, J. C., Kang, M., Son, M. J., Kim, Y. S., & Woo, S. H. (2011). Voltagedependent anion channel 2 modulates resting Ca2 + sparks, but not action potentialinduced Ca2 + signaling in cardiac myocytes. *Cell Calcium*, 49(2), 136–143. https:// doi.org/10.1016/j.ceca.2010.12.004.
- Sun, G., Ukkola, O., Rankinen, T., Joanisse, D. R., & Bouchard, C. (2002). Skeletal muscle characteristics predict body fat gain in response to overfeeding in never-obese young men. *Metabolism, Clinical and Experimental*, 51(4), 451–456. https://doi.org/10. 1053/meta.2002.31324.
- Takeda, K., Sriram, S., Chan, X. H. D., Ong, W. K., Yeo, C. R., Tan, B., ... Sugii, S. (2016). Retinoic acid mediates visceral-specific adipogenic defects of human adipose-derived stem cells. *Diabetes*, 65(5), 1164–1178. https://doi.org/10.2337/db15-1315.
- Thornton, K. J., Chapalamadugu, K. C., Eldredge, E. M., & Murdoch, G. K. (2017). Analysis of Longissimus thoracis protein expression associated with variation in carcass quality grade and marbling of beef cattle raised in the Pacific Northwestern United States. Journal of Agricultural and Food Chemistry, 65(7), 1434–1442. https://doi.org/ 10.1021/acs.jafc.6b02795.
- Traaseth, N., Elfering, S., Solien, J., Haynes, V., & Giulivi, C. (2004). Role of calcium signaling in the activation of mitochondrial nitric oxide synthase and citric acid cycle. *Biochimica et Biophysica Acta - Bioenergetics*, 1658(1–2), 64–71. https://doi.org/10. 1016/j.bbabio.2004.04.015.
- Wang, B., Yang, Q., Harris, C. L., Nelson, M. L., & Busboom, J. R. (2017). Of retinoic acid : A review, (509). 100–106. https://doi.org/10.1016/j.meatsci.2016.04.003. Nutrigenomic.
- Xu, X., Li, J., Sun, X., Guo, Y., Chu, D., Wei, L., ... Shen, L. (2015). Tumor suppressor NDRG2 inhibits glycolysis and glutaminolysis in colorectal cancer cells by repressing c-Myc expression. Oncotarget, 6(28), 26161–26176. https://doi.org/10.18632/ oncotarget.4544.
- Zhang, Q., Lee, H. G., Han, J. A., Kim, E. B., Kang, S. K., Yin, J., ... Choi, Y. J. (2010). Differentially expressed proteins during fat accumulation in bovine skeletal muscle. *Meat Science*, 86(3), 814–820. https://doi.org/10.1016/j.meatsci.2010.07.002.
- Zhang, Y., Zan, L., Wang, H., Qing, L., Wu, K., Quan, S., ... Wang, C. (2011). Differentially expressed genes in skeletal muscle tissues from castrated Qinchuan cattle males compared with those from intact males. 135, 76–83. https://doi.org/10.1016/j.livsci. 2010.06.070.

Points clés de l'article 3:

Molecular signatures of muscle growth and composition deciphered by the meta-analysis of age-related public transcriptomics data

14 OBJECTIFS

- Identifier des ARNm dont l'abondance musculaire des ARNm varie **selon l'âge des bovins**, et susceptibles de signer des différences d'adiposité musculaire.
- Réutiliser et ré-analyser par méta-analyse les données transcriptomiques générées sur des muscles bovins.
- Mettre en place les méthodologies liées à l'identification, la sélection, l'agrégation, la mise à jour et l'analyse de grands jeux de données transcriptomiques.



15 DEMARCHE EXPERIMENTALE/ANALYTIQUE

Figure 28 : Schéma de sélection et d'analyse des données publiques transcriptomiques issues du muscle bovin

Critère de sélection utilisé pour le facteur de variation âge : suppression des données relatives à l'âge fœtal (n=6), des données provenant de RNAseq (n=1) et des données dupliquées (n=3). DE : différentiellement exprimés.

Cinq races pures ou croisées ont été utilisées pour les analyses transcriptomiques : Angus, Wagyu x Hereford, Piedmontese x Hereford, Belmont Red et Chinese Red Steppe

L'âge des animaux est compris entre 30.5 jours et 732 jours. Afin de rendre les races comparables, un âge physiologique a été calculé en prenant en compte le poids de l'animal à l'abattage divisé par le poids final adulte pour chaque race. Lors de l'interprétation des résultats ceci permet de « gommer » l'effet race pour une meilleure comparaison des études.

16 CONCLUSIONS

Identification de :

- 715 gènes communs aux 10 études rapportant un transcriptome différentiel selon l'âge
- 238 gènes différentiellement exprimés entre 2 âges
- 97 gènes différentiellement exprimés dans au moins 2 jeux de données et la métaanalyse
- 17 gènes différentiellement exprimés uniquement dans la méta-analyse
- 3 clusters d'expression identifié via la méthode des k-means, regroupant respectivement 78, 24 et 136 gènes différentiellement exprimés
- Les gènes liés aux métabolismes glycolytiques (GPI, PDHA1, PGAM2, PGK1 et PKM) et oxydatifs (IDH3B, MDH2, OGDH, SDHA et SUCLG1) sont régulés en fonction de l'âge des bovins. De plus, MYH1, lié aux fibres musculaires glycolytiques, et TNNT1, liés aux fibres oxydatives sont régulées dans le même sens.
- Les gènes MYOG et MAPK14, des facteurs de transcription clés du muscle, ainsi que TXNIP pourraient contribuer à la dynamique de régulation des métabolismes glycolytiques et oxydatifs.
- Plusieurs gènes dont l'abondance augmente avec l'âge des bovins sont liés au métabolisme des lipides : APOE, LDLR, MXRA8, HSP90AA1 et FABP4
- Enfin, d'autres gènes MAPK14, YBX1, YWAH et ERBB2 pourraient réguler la différenciation de cellules progénitrices du muscle vers le lignage adipeux.

Points clés – Article 3

Molecular signatures of muscle growth and composition deciphered by the meta-analysis of age-related public transcriptomics data Jeanne BAZILE¹, Florence JAFFREZIC², Patrice DEHAIS^{3,4}, Matthieu REICHSTADT¹, Christophe KLOPP^{3,4}, Denis LALOE², Muriel BONNET¹

- 8 1 INRAE, UMR Herbivores, Université Clermont Auvergne, VetAgro Sup, F-63122 Saint-Genès-
- 9 Champanelle, France
- 10 2 INRAE, UMR1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France
- 11 3 Plate-forme bio-informatique Genotoul, Mathématiques et Informatique Appliquées de
- 12 Toulouse, INRAE, Castanet Tolosan, France.
- 13 4 SIGENAE, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan, France.

14 Keywords:

15 meta-analysis, public data, muscle, bovine

16 Abbreviated title:

17 Meta-analysis of age-related muscle transcriptomics data

18 Address for reprint requests and other correspondence:

- 19 M.Bonnet, INRAE, UMR Herbivores, Université Clermont Auvergne, VetAgro Sup, F-63122
- 20 Saint-Genès-Champanelle, France (e-mail: Muriel.bonnet@inrae.fr)
- 21 Phone number : +33 4 73 62 47 01

22 **ORCID**:

- 23 Jeanne BAZILE : 0000-0001-5920-5279
- 24 Florence JAFFREZIC : 0000-0001-7579-6419
- 25 Christophe KLOPP : 0000-0001-7126-5477

26 Denis LALOE : 0000-0001-8359-0760

27 Muriel BONNET : 0000-0001-7193-3543

28 Abstract

29 The lean-to-fat ratio is a major issue in beef meat industry both for carcass and meat production perspectives. This industrial perspective has motivated meat physiologists to use 30 31 transcriptomic to decipher mechanisms behind fat deposition within muscle during the time 32 course of muscle growth. However, a synthetic biological information from this volume of data 33 remains to be produced to identify mechanisms found in various breeds and rearing practices. 34 We conducted a meta-analysis on 10 transcriptomic datasets stored in GEO database, from 35 the Longissimus thoracis of five different bovine breeds divergent by the age. We updated 36 gene identifiers on the last version of the bovine genome (UCD1.2) and the 715 genes common 37 to the 10 studies were subjected to the meta-analysis. Of the 238 genes differentially 38 expressed (DEG), we identified a transcriptional signature of the dynamic regulation of 39 glycolytic and oxidative metabolisms that agrees with a known shift between those two 40 pathways from the animal puberty. We proposed some master genes of the myogenesis, 41 namely MYOG and MAPK14, as probable regulators of the glycolytic and oxidative 42 metabolisms. We also identified up-regulated genes related to lipid metabolism (APOE, LDLR, 43 MXRA8 and HSP90AA1) that may contribute to the expected enhanced marbling as the age 44 increases. Lastly, we proposed transcriptional signature related to the induction (YBX1) or 45 repression (MAPK14, YWAH, ERBB2) of the commitment of myogenic progenitors into the 46 adipogenic lineage. The next relationship analysis between the abundance of the identified 47 mRNA and marbling values remains to be analyzed in a marbling biomarkers discovery 48 perspectives.

49 Introduction:

The management of muscle mass and/or of meat eating qualities goes through a well understanding of the molecular drivers of muscle growth and composition. Indeed, in cattle as in others vertebrates, muscle is composed of several tissues, such as muscle fibers as well as connective and adipose tissues (44). Genetic (breed, gender...) and non-genetic (age, nutrition...) factors strongly influence the growth of each cell type within muscle (10). For example, the increase in age, the most influent non-genetic factor, decreases the proportion 56 of muscular fibers while the proportion of intramuscular adipose and connective tissues 57 increases (10, 28, 44, 51). Consequently, these age-related variations in muscular composition 58 affect the major sensorial and nutritional meat-eating qualities (44), which has motivated age-59 dependent high throughput molecular characterization of bovine muscle (5, 16, 32, 35, 46, 53, 60 58, 60, 75). The aim of these studies was to propose indicators and knowledge to manage an 61 optimal trade-off between muscle growth and composition for economic and consumers 62 satisfactory purposes of beef production. However, to date, a small consensus on molecular 63 signature were repeated such as PPARy signaling from mRNA (5) or glycolytic and oxidative 64 pathways for proteins (15, 17, 20, 23, 66). This few consensus may result from the difficulty to merge and analyze together huge amount of data provided by high throughput methods, as 65 66 well as the confounded effect of the breed, gender, nutrition on the age-related molecular 67 signatures. Meta-analysis (55), as well as the emerging concept of data re-use for 68 transcriptomic (micro-array or RNAseq (59)) or proteomic results (74), offers the opportunity 69 to produce knowledge from available data with several advantages : reduced cost, reduced 70 use of animals, reliable and robust results as they were recorded in several independent 71 studies. The first and rare attempts of meta-analyses of bovine public "omics" data have 72 successively provided the signatures of the lactation process in dairy cattle (25), predictors of 73 oocyte quality in bovine (37), and conserved gene expression patterns between species for 74 normal and pathological muscles (7). We thus hypothesized that the meta-analysis of available 75 muscular transcriptomics data would allow identifying robust molecular signatures of muscle 76 growth and composition, especially marbling. We use the variation of post-natal ages across 77 studies as a proxy of the physiological variation in muscular mass, composition and 78 intramuscular lipid content. We use, reuse, re-annotated and analyzed 10 muscular 79 transcriptomic datasets available from meat producing bovines to describe the transcriptomic 80 signature of muscle according to muscle growth and whatever the breeds. To our knowledge, 81 this is the first attempt to reuse the publicly available bovine muscular transcriptomic data to 82 identify conserved patterns between breeds and rearing practices, and to propose potential 83 indicators of muscular growth and composition in bovine.

84

Muscle mass and/or meat eating qualities management needs a good understanding of muscle growth and composition molecular drivers. Indeed, in cattle as in others vertebrates,

87 muscle is composed of several tissues, such as muscle fibers as well as connective and adipose 88 tissues (44). Genetic (breed, gender...) and non-genetic (age, nutrition...) factors strongly 89 influence the growth of each cell type within muscle (10). For example, aging, the most 90 influent non-genetic factor, decreases muscular fiber and increases intramuscular adipose 91 and connective tissues proportions (10, 28, 44, 51). Consequently, these age-related muscular 92 composition variations affect sensorial and nutritional meat-eating qualities (44). This has 93 motivated age-dependent high throughput molecular characterization of bovine muscle (5, 94 16, 32, 35, 46, 53, 58, 60, 76). The aim of these studies was to propose indicators and increase 95 knowledge in order to achieve an optimal trade-off between muscle growth and composition 96 for economic and consumer satisfaction purposes. However, to date, few consensus on the molecular signature of the growth of muscle fibers relatively to connective and adipose tissues 97 98 have been proposed (17, 20, 23, 66). This may come from the difficulty to merge and analyze 99 huge amounts of data provided by high throughput methods, as well as the confounded 100 breed, gender and nutrition effects on age-related molecular signatures. Meta-analysis (55), 101 as well as the emerging concept of data re-use for transcriptomic (micro-array or RNAseq (59)) 102 or proteomic results (74), offers the opportunity to produce knowledge from available data 103 with several advantages : reduced cost, reduced use of animals, reliable and robust results as 104 they were recorded in several independent studies. The first and rare attempts of bovine 105 public "omics" data meta-analyses have successively provided signatures of lactation process 106 in dairy cattle (25), predictors of oocyte quality in bovine, and conserved gene expression 107 patterns between species for normal and pathological muscles (7). We thus hypothesized that 108 the meta-analysis of available muscular transcriptomic data would allow identifying robust 109 molecular signatures of muscle growth and composition, especially marbling. We use the 110 variation of post-natal ages across studies as a proxy of physiological variation in muscular 111 mass, composition and intramuscular lipid content. We use, reuse, re-annotated and analyzed 112 10 muscular transcriptomic datasets available from meat producing bovines to describe 113 transcriptomic signatures of muscle according to muscle growth and breed. To our knowledge, 114 this is the first attempt to reuse publicly available bovine muscular transcriptomic data to 115 identify conserved patterns between breeds and rearing practices, and to propose potential 116 indicators of muscular growth and composition in bovine.

117

118 Material & Methods

119 The work presented in this paper was conducted in accordance with guidelines previously 120 proposed for meta-analysis of microarrays datasets (55, 73). All data analyses were performed 121 using R statistical software (version 3.5.2)(54).

122

123 1. Data retrieval and expression as ratio

124 Datasets and their corresponding raw expression data from bovine, post-natal and age-related 125 muscle, were collected from the public repositories Gene Expression Omnibus (GEO, 126 http://www.ncbi.nlm.nih.gov/geo/)(8) or Pubmed (<u>https://www.ncbi.nlm.nih.gov/pubmed</u>) 127 of the National Center for Biotechnology Information (NCBI) at the National Institutes of 128 Health (NIH), as well as from the Web of Science (WoS; https://www.webofknowledge.com/). 129 These publication (46, 53) and datasets (from GSE5659, GSE48136, GSE21782) were identified 130 using keywords related to muscle (muscle, myo*, satellite cells), growth (growth, hypertrophy, 131 atrophy, hypermuscled, hypermuscularity, double-muscled, marbling), bovine specie (bovine, 132 calf, steers, heifers, cows, beefs, Bulls), methods (transcript*, microarray, chip, gene array) 133 and exclusion criteria (fetal, disease, cardio*, heart). Of the hundred of datasets retrieved, we 134 selected 25 transcriptomic datasets (18 from GEO, 7 from the WOS, see the PRISMA flowchart 135 on Figure 1), obtained in the most studied bovine *longissimus thoracis* muscle, and they 136 reported comparison between bovine differing within the study only by the age (similar breed, 137 rearing practice and sex). We removed duplicate data from the GSM132170 and GSM132194 138 (41) (GSM stands for GEO Sample and corresponds to 1 dataset, a GSE (GEO Series Experiment) 139 may comprise many GSM) that have been re-analyzed in the GSE5659 and were more recently 140 updated. We then excluded a dataset that has been obtained by RT-PCR, as there would be 141 too few genes to include in the meta-analysis (40), a RNA-seq dataset (31), and a micro-array 142 dataset due to the lack of gene identifiers (GSE60844). Six datasets comparing a fetal and a post-natal ages were excluded because cellular and biological pathways involved in pre- and 143 144 post-natal muscle growth are very different (50). Lastly, 3 datasets were excluded because the 145 lack of sequences linked to the identified gene did not allow to update the gene identifiers on the current version of the bovine genome. Thus, 10 datasets were subjected to data 146 147 preprocessing, gene identifiers update and meta-analysis as described thereafter (Figure 1). 148 These 10 datasets were stored in GEO as 3 GSE, namely GSE21782 (1 dataset), GSE48136 (3

149 datasets) and GSE5659 (6 datasets), and reported mRNA abundance assayed by microarrays 150 in muscles from 5 bovine pure- or cross-breeds aged from 31 to 732 days after birth (Table 1). 151 The GSE5659 reported data from 147 samples differing by the breed, the nutrition or the age. 152 Thus, 6 sub-datasets were created from the GSE5659 (with 2 to 4 technical replicates per 153 animal) to compare 2 ages. These sub-datasets were named E2, E3, E4, E6, E8 and E9 and were 154 produced from a non-commercial microarray (42). We have created one sub-dataset from 155 GSE21782, namely the study E5, with data from 3 muscles produced from the Affymetrix 156 Bovine Genome Array (53). Three sub-datasets were created from GSE48136, named E1, E7 157 and E10 with data from 13 to 14 muscles. The micro-array assays were performed on non-158 commercial microarray (46).

| STUDY N° | GSE N° | Age of the older animals (days) | Age of the younger animals (days) | Sex | Breed | Number of animals used for each age | Live body weight for older animals (kg) | Live body weight for younger animals (kg) | Theorical adult weight (kg) | Physiological age for older animals | Physiological age for younger animals | Maturity | Number of genes analysed after the ID update |
|-------------|-----------|---|---|-----|---------------------------|--|---|---|--------------------------------------|---|--|----------|---|
| 1 | 48136 | 215 | 155 | C | Angus | 7 | 265 | 200 | 900 | 0,29 | 0,22 | early | 7054 |
| 2 | 5659 | 366 | 244 | C | Belmont Red | 9 | 325 | 250 | 950 | 0,34 | 0,26 | late | 999 |
| 3 | 5659 | 214 | 92 | F | Piedmontese x Hereford | 9 | 186,2 | 93,6 | 533 | 0,35 | 0,18 | inter | 999 |
| 4 | 5659 | 214 | 92 | F | Wagyu x Hereford | 9 | 185,9 | 92,1 | 510 | 0,36 | 0,18 | early | 999 |
| 5 | 21782 | 732 | 31 | Μ | Chinese Red Steppe | 8 | NS | NS | 1000 | 0,1 | 0,4 | early | 10638 |
| 6 | 5659 | 458 | 366 | C | Belmont Red | 9 | 400 | 325 | 950 | 0,42 | 0,34 | late | 999 |
| 7 | 48136 | 275 | 155 | C | Angus | 7 | 387 | 200 | 900 | 0,43 | 0,22 | early | 6993 |
| 8 | 5659 | 366 | 214 | F | Piedmontese x Hereford | 9 | 269,3 | 186,2 | 533 | 0,51 | 0,35 | inter | 999 |
| 9 | 5659 | 366 | 214 | F | Wagyu x Hereford | 9 | 267,5 | 185,9 | 510 | 0,52 | 0,36 | early | 995 |
| 10 | 48136 | 375 | 155 | C | Angus | 7 | 560 | 200 | 900 | 0,62 | 0,22 | early | 6996 |

The comparison between 2 ages from a unique GSE were numbered, and comparison concerned data from the older bovine divided by the data from the younger one. In the column sex, C=Castrated, F=Females and M=Male. The physiological age is a value between 0 and 1, calculated as the live body weight at slaughter divided by the assumed adult weight. This physiological age is an indicator of the animal maturity. The type of bovine were described by their rearing purpose as meat production (beef), milk production or dual (meat and milk production) and by their maturity: either early maturing breed (early), late maturing breed (late) or intermediate (inter). The number of genes remained after ID updates were subjected to meta-analysis.





159

161 Figure 1 : Prisma flowchart of the data selection. In total, 10 studies comparing muscular transcript abundance

- 162 between 2 ages of bovine, were included in the meta-analysis. The characteristics of the included studies were 163 given in Table 1. N is the number of datasets/publications selected.
- 164

8

165 2. Data re-normalization

166 Most of the time, raw data from GEO correspond to already pre-processed and/or normalized 167 data. However, to remove artefacts and to ensure that each study follows the same procedure 168 we normalized the pre-normalized microarray data. To achieve this goal, each of the selected 169 microarray dataset was verified manually to take into account if they were labelled according 170 to a dye swap or a dye switch protocol, if the data were paired, normalized and log-171 transformed. According to these indications and mainly for data from GSE21782 and 172 GSE48136, the values were normalized with the anapuce package normalisation function (77), 173 converted into log base 2 and the sign was reversed in case of a dye swap or dye switch. The 174 average of abundances of replicated probes or of technical replicates were calculated. As data 175 from GSE5659 (6 datasets), and thus most of half of the available data, was expressed as ratio, 176 for the remaining 4 datasets we calculated ratio per mRNA as the log abundance in the muscle 177 of the oldest bovine divided by the log abundance in the youngest animal. Based on the 178 similarity of variance abundances between the biological replicate, 59 biological replicates 179 categorized within 10 datasets were included in the meta-analysis.

180 3. Gene identifiers update to match probes with the last version of bovine181 genome

As the datasets were obtained over the last decade and were performed on different probes, 182 183 we had to update the genes targeted by those probes. Thus, probe IDs were matched with 184 their corresponding official gene symbol updated on the current bovine genome (UCD1.2) 185 using the pipeline sigReannot as previously described by Casel et al. (13). Briefly, a BLAST 186 (blastall 2.2.26) (3) was performed using the FASTA files provided within the raw data GPL file 187 of each microarray. The criteria applied for successful blast results were continuous stretch of 188 at least 15 base pairs for 50 or 20 bp for 70 mers and a global identity percentage of at least 189 85%. The sequences of probes were blasted against various banks of the current bovine 190 genome, namely verified bovine transcripts, bovine ncRNA, against the whole genome, and 191 against annex databases namely RefSeq Bos Taurus, RefSeq Sus scrofa, RefSeq Mus musculus, 192 RefSeq Homo sapiens and ab initio bovine cDNA (all versions available in June 2018). The 193 results were stored into a relational database in order to ease cross data comparisons and 194 final gene assignations. The success rate of gene name update was at least of 23.7% (E6) and 195 up to 64.7% (E5) of probes IDs updated with a current gene name.

196 4. Meta-analysis

197 Since the number of genes between studies were different (from 995 to 10638 gene names), 198 the 715 common genes across the 10 datasets were gathered together to make the merged 199 dataset loaded in the meta-analysis. The meta-analysis was performed on the ratio values 200 from the 10 selected studies based on P value combination using the inverse normal method. 201 Meta-analysis calculates for each study, the p-value of differential ratio for paired data with 202 moderated t-test method (Limma) using the package metaMA (version 3.1.2) (45). P-values 203 were corrected for multiple testing using a Benjamini-Hochberg correction and were 204 considered as significant when P < 0.05. Whenever a gene was found several times in a same 205 sample, the mean of the abundance ratios for this gene was calculated and used for the meta-206 analysis. By the end, meta-analysis was performed for 715 unique genes that had expression 207 values across all the 10 studies. Based on the Benjamini-Hochberg p-values (< 0.05) a list of 208 differentially abundant mRNA between 2 post-natal ages was identified.

209

5. Cluster Analysis and Annotation according to Gene Ontology 210 Clusters of co-expressed genes were identified using an unsupervised k-means procedure 211 212 implemented in the R package pheatmap (version 1.0.12(38)) with the option kmeans k, and 213 using the ratio values of the differentially abundant mRNA between 2 post-natal ages. Using 214 this threshold, all the 238 genes listed as differentially expressed by the meta-analysis were 215 submitted to the k-means clustering. The number of clusters was defined with the best.cutree 216 function (https://github.com/larmarange/JLutils/blob/master/man/best.cutree.Rd) ran on R 217 software. Data were visualized with a heatmap produced with the pheatmap function in 218 pheatmap package (38). The figures were handled with Inkscape (2).

Genes from each cluster were analyzed for Gene Ontology (GO) enrichment using ProteINSIDE (http://www.proteinside.org/,(34)). To take advantage of the better annotations within human, the GO enrichment analysis was performed in the human species, with a threshold of 0.05 for the Fisher's exact test performed on GO terms and adjusted with Benjamin-Hochberg method. The secreted proteins and potential QTL were predicted using ProteInside as well, using the bovine species to perform the QTL prediction.

225 6. Functional analyses

226 The topological analysis of networks constructed between the gene products of the DEG and 227 the human proteins (as protein-protein interactions were much more described in human 228 than bovine) declared in Psicquic (4) was realized with ProteINSIDE (34). We recorded the 229 number of interactions, as well as centralities that were proven as efficient to reveal proteins 230 that play important roles in a network, namely betweeness and closeness centralities. Briefly, 231 betweeness centrality quantifies how frequently a node is on the shortest path between every 232 pair of nodes for detecting bottlenecks in a network. Closeness centrality quantifies how short 233 are minimal paths from a given node to all others, a large closeness indicates that a node is 234 close to the topological center of the network. Centralities were calculated on a PPI network 235 that linked the gene products identified as differentially abundant (DEG) in the meta-analysis 236 and interacting proteins declared in IntAct and UniProt databases. Only PPI that were reported 237 thanks to experiments were considered, thus 6903 PPI were identified between 3249 proteins 238 of these 214 were DEG from the meta-analysis (version of ProteINSIDE Database was 1.2.11 239 and of Psicquic 1.4.11).

240 7. Analysis of genes mapped in QTL regions

246

The genes that were found as differentially expressed between 2 ages were mapped against QTL regions using ProteINSIDE webservice that uses resources from the Cattle QTL database (<u>https://www.animalgenome.org/cgi-bin/QTLdb/BT/index</u>). We mapped genes in regions where QTL related to muscle growth (Lean meat yield, Longissimus muscle area), intramuscular fat content (Intramuscular fat, Marbling score) and carcass adiposity (Fat

thickness at the 12th rib, Subcutaneous fat, Rib Fat) were reported.

247 8. Mining of data according to the physiological age of bovine 248 Across lifespan, the rate of hypertrophy and growth of *Longissimus thoracis* muscle, as well as 249 the percent of intramuscular fat are very different among bovine breeds (28) because of 250 differences in adult weight. To ensure a relevant datamining across the 5 pure- or cross breeds 251 under the meta-analysis we calculated the degree of maturity (DoM; Figure 2) of the animals that is the ratio between the slaughter body weight (kg) and the adult body weight (kg). The 252 253 10 studies used for the meta-analysis were categorized according to the DoM. The first group 254 is composed by studies E1, E2, E3 and E4 that compared muscles from bovine from 0.18 up to 255 0.36 of adult weight or DoM. The second group is composed by studies E5, E6, E7, E8 and E9 256 that compared bovines from 0.1 to 0.52 of DoM. The third group with the study E10 that

reported results from bovine muscles from 0.22 to 0.62 of DoM. Regarding the muscle composition, the amount of intramuscular lipids strongly increased from 8 to 14 months of age in early and late maturing breeds (28) corresponding to a DoM of 0.30 +/- 0.07. An average ratio of mRNA abundance per group of DoM was calculated as the means of the individual abundance ratio between old and young bovine per studies and by group.



262

Figure 2 : Groups of physiological age. Each study is colored according to the group it belongs to: blue for group1, orange for group2 and green for group3. Study E5 is in dot line as the weight of the animals was not specified.E1, E7 and E10 are superposed. The first group is composed by studies E1, E2, E3 and E4 that compared muscles from bovine from 0.18 up to 0.36 of adult weight or DoM. The second group is composed by studies E5, E6, E7, E8 and E9 that compared bovines from 0.1 to 0.52 of DoM. The third group with the study E10 that reported results from bovine muscles from 0.22 to 0.62 of DoM.

269 **Results**

1. Identification of Differentially Expressed Genes 270 271 We considered 125 bovines from Angus, Belmont Red, Chinese Red Steppe breeds as well as 272 from Piedmontese x Hereford and Wagyu x Hereford crossbreds over the time course of LT 273 muscle growth, i.e. from 31 to 732 days of post-natal life (Figure 2). A total of 10 micro-array 274 datasets have been used, which provided 715 genes that were quantified for their mRNA 275 abundance in all of the datasets. Of these, 238 genes were identified as differentially 276 expressed (DEG, p-value adjusted by the Benjamini-Hochberg correction < 0.05) between 2 277 ages and whatever the breed. Of these 214 mRNA were over abundant and 24 were less

abundant when older bovines were compared to the youngest animals. The top 20 mRNA with
the highest significant (P <0.05) fold-change were FMO5, CYCS, TCAP, SMG7, RPL29, CTSS,
PGAM2, STRA6, CYC1, STXBP1, ASPH, MAPK14, GPRC5C, ARPC5, DUSP1, RPS12, ACSL3, TIMP1
and HSPA5 with a ratio higher than 2.5 (log 2 fold change higher than 1.35 when old bovines
were compared to the youngest ones).

283

284 Among the 238 DEGs, 97 mRNA were identified as differentially abundant between two ages 285 in at least two studies and the meta-analysis. Of these, 15 DEGs were identified in more than 286 half of the studies and the meta-analysis, namely MB was DEG in six studies and the meta-287 analysis, ASB12, FN1, LRRC20, PKM in five studies and the meta-analysis, and ACTN3, ATF4, 288 CLNS1A, CLU, COPG1, CYC1, GPI, PFKFB3, PGAM2, PGK1 four studies and the meta-analysis 289 Among the 238 DEGs between two ages, 17 were exclusively identified in the meta-analysis, 290 namely APOE, CAMTA2, CTSH, HMBS, HSP90AA1, LDLR, MGC148714, MXRA8, MYL12A, NNT, 291 RPL32, RPLPO, RPS13, SGCE, SOD2, STAC3 and UQCR11.

292 2. Cluster analysis and Gene ontology annotation

To group DEG with similar expression, we applied a k-means clustering to the 238 mRNA abundance. The k-means result (Figure 3) produced three distinct clusters, whose mean profiles are given in Figure 4. They were grouped into two global profiles first consisting of upregulated expression (clusters 1 and 3) and down-regulated (cluster 2) expression when older animals were compared to younger ones.



298

Figure 3 : Clusters of 238 DEGs from the meta-analysis. The clustering was made using the k-means method and heatmap was produced accordingly to this clustering using R package pheatmap. Gene ontology enriched terms identification. The results obtained with ProteINSIDE and enrichments were expressed as -log10 (pvalue) to visually plot them on graphs, which means that a -log10 (P value) of 3, 2, and 1.3 correspond to a p-

value of 0.001, 0.01, and 0.05, respectively. The genes or the number of genes within the GO annotation werewritten next to the bar.



Figure 4: Mean profiles of log ratio of the gene expression by cluster. The number in abscissa corresponds to the number of the study (1=E1, 2=E2,...,10=E10).

308

305

309

310 Cluster 1 contained 78 genes with the highest abundances in studies 10, 5 and 3 comparing 311 Angus at 375 versus 155 days, Chinese Red Steppe at 732 versus 31 days and Angus at 215 312 versus 155 days. The lowest abundance were recorded in studies 8 and 9 comparing 313 Piedmontese x Herford and Waguy x Hereford at 366 versus 214 days. The top biological 314 processes related to 77 genes that were annotated by enriched (P< 1.13E-03) GO terms are 315 reported in Figure 3. Of the 77 annotated genes, ACTN3, TNNT1, TNNC2, MYBPC1, TPM2, 316 MYL3, STAC3, MB and MYH1 were annotated by 5 terms related to muscle organization and 317 contraction (muscle filament sliding, skeletal muscle contraction, slow-twitch skeletal muscle 318 fiber contraction, muscle contraction, transition between fast and slow fiber, sarcomere 319 organization and skeletal muscle cell differentiation). PGK1, LDHA, OGDH, PKM, GPI, MDH2, 320 ATF4, GOT2, PGK1, PDHA1, BSG, and PLN were annotated by 6 GO terms related to glucose 321 metabolism and signaling (glycolytic process, gluconeogenesis, canonical glycolysis, pyruvate 322 metabolic process, response to insulin). DHRS3, NNT, MDH2, PDHA1, OGDH, CYC1, SOD2, 323 LDHA, ATP5F1B, COX7B, ATP2A1, GOT2 were annotated by 7 terms related to oxidative 324 metabolism (oxidation-reduction process, tricarboxylic acid cycle, proton transmembrane 325 transport, oxaloacetate metabolic process, 2-oxoglutarate metabolic process, mitochondrial 326 ATP synthesis coupled proton transport and ATP biosynthetic process). Less numerous DEG

15

were annotated by terms related to cell volume homeostasis (CLNS1A, AQP1) skeletal muscle
(FOS, ASB2) and fat cell (MB, FABP4) differentiation, as well as negative (ANGPTL4, NNT, AQP1,
CD74) and positive (LDHA, TXNIP, ATF4, CTSD) regulation of apoptotic process.

330 Cluster 2 contained 24 genes with the highest abundance ratios in studies 4 and 9 comparing 331 Waguy x Hereford at 214 versus 92 days and at 366 versus 214 days. The lowest abundance 332 ratios were recorded in studies 1 and 10 comparing Angus at 215 versus 155 and at 375 versus 333 155 days. The top biological processes related to the 24 genes that were annotated by 334 enriched (P< 3.69E-03) GO terms are reported in Figure 3. Of these 24 genes, RPS9, RPL18, 335 RPL32, RPLP0, RPS13, IGFBP3, ATG5, RBBP7 were annotated by terms related to the 336 translation (translational initiation, translation, cytoplasmic translation, post-translational 337 protein modification). RBBP7, HYAL2, ENO1, H3F3A were annotated by GO terms related to 338 cell cycle (negative regulation of cell growth, negative regulation of chromosome condensation, telomeric heterochromatin assembly, negative regulation of G0 to G1 339 340 transition). IGFBP3, CYP19A1, RPLP0 were annotated by terms related to hormone or cytokine 341 signaling (positive regulation of insulin-like growth factor receptor signaling pathway, positive 342 regulation of estradiol secretion, interleukin-12-mediated signaling pathway).

343 Cluster 3 contained 136 DEG with the highest positive abundance ratios in the studies 1, 7 and 344 10 comparing Angus at 215 versus 155, at 155 versus 275 and at 375 versus 155 days of life. 345 The lowest negative abundance ratios were recorded in studies 8 and 2 comparing 346 Piedmontese x Hereford at 366 versus 214 days and Belmont Red at 366 versus 244 days of 347 life. The top biological processes related to the 132 genes that were annotated by enriched 348 (P< 1.61E-6) GO terms are reported in Figure 3. MAPRE2, MYOG, DDIT3, GNAI1, USP16, 349 PIK3C3, ENSA, TERF1, NR3C1, ANAPC11, DUSP1, CSF2, CTSH, CXCL12, FN1, CAPNS1, NPM1, 350 TIMP1, ERBB2, HAS2, SFRP4, PHB, ASPH, TIMP2 were annotated by GO terms related to cell 351 cycle, and positive or negative regulation of cell population proliferation. FSTL1, TNF, CSRP3, 352 PMCH, SFRP4, MGP, BIN1, CREG1, MYOG, MAPK14, ELN, CSRP3, SGCE, UTRN were annotated 353 by GO terms related to muscle development (multicellular organism development, skeletal 354 muscle tissue development, muscle organ development). ADRB2, PCBP1, MYOG, TNF, NR3C1, 355 YBX1, CAMTA2, NFIC, SERPINE1, MAPK14, NFKB2, DDIT3, NPM1, USP16, CSRP3, SMG7, TERF1, 356 HSP90AA1, APOE, PHB, ASPH, YWHAH, NCOR1, YBX3, CREG1, CTDSP2 and ERBB2 were 357 annotated by GO terms related to the regulation of transcription (positive regulation of 358 transcription by RNA polymerase II, telomere maintenance via telomerase, positive regulation 359 of transcription, DNA-templated, negative regulation of transcription by RNA polymerase II, 360 regulation of transcription by RNA polymerase II). GSPT1, RPS2, MRPL2, RPL19, EIF4G2, RPL29, 361 RPL27A, MRPL35, RPL15, RPS12, RPL24, RPS12, HSPA14, CST3, APOE, FSTL1, TIMP1, FN1, 362 MXRA8 were annotated by GO terms related to the regulation of translation (translation, 363 translational initiation, cytoplasmic translation, post-translational protein modification). 364 Other biological processes relevant to muscle structure and metabolism were enriched within 365 the DEG from cluster 3: oxidation-reduction process (NDUFB9, IFI30, CYCS, SDHA, FMO5, 366 ACOX2, UQCR11, ACADS, ASPH, IDH3B), lipid metabolic process (ACADS, FDFT1, ACOX2, 367 ACSL3, LDLR, LIPE, PLA2G2A, APOE, PLA2G12A), muscle contraction (TPM1, FXYD1, MYL12A, 368 ASPH, UTRN) and extracellular matrix disassembly (TIMP2, CTSK, CAPNS1, TIMP1, CTSS).

369 3. Data mining

370

371 Identification of genes with the highest regulatory potential through network analysis

In order to identify the genes with the highest regulatory potential we evaluated the closeness and betweeness centralities of the DEG gene product within a PPI network. Of the 50 top genes with the highest closeness and betweeness centralities, 29 were found in the two lists of centralities. Of these YBX1, HSPA5, STXBP1, YWHAH, SDHA, HSP90AA1 and ERBB2 were among the genes with the highest fold change. Six additional DEG (RPL29, CYCS, MAPK14, FDFT1, MYOG, AQP1) were among the genes with the highest fold change, and closeness or betweeness centralities (Figure 5).



379

Figure 5: Venn diagram (6) highlighting genes among the 238 DEG of the meta-analysis that were found with the highest betweeness centrality (BC), Closeness centrality (CC) and Fold change (FC).

382

383 Identification of QTL related to muscle growth and marbling around DEG identified by meta 384 analysis

385 Of the 238 DEG genes identified by the meta-analysis, OGDH, EIF4A2, ANGPTL4, RBM38, 386 ROMO1, CYC1, AQP1, SLA, DES, IGFBP3 were near QTL for marbling fat or marbling score. 387 Moreover CLNS1A, PKM, ACAT1, COQ8A, ATG101, CSDE1, CLU, GOT2, TXNIP, LDHA, CAMK2D, 388 ANGPTL4, HMBS, CYC1, IRF8, SBDS, MYH1, PDLIM1, SLA, LRRC20, ATP2A1, VEGFC were near 389 QTL for lean meat yield. STAC3, CKM, OGDH, ATG101, GPI, TPPP3, EIF4A2, NNT, AQP1, RTN2, 390 DES, ATG5, IGFBP3 were shown near QTL related to *Longissimus muscle* area. Of these, GOT2, 391 TXNIP (Longissimus muscle area), CYC1 (marbling and lean meat yield QTL) and AQP1 392 (marbling and Longissimus muscle area) were among the top 50 DEG with the highest fold 393 changes.

394 Identification of the most differentially abundant mRNA regarding the phases of muscular 395 growth

396 Gene with ratio of abundance between 2 ages (not log2 ratio) below 0.5 and higher than 2 397 were mined according to the growth phase of the bovine. We listed mRNA with the top fold 398 change during the early (up to 0.32 of final body weight also called degree of maturity DoM, 399 group 1), mid (up to 0.52 DoM, group 2) and more than the mid (up to 0.62 DoM, group 3) 400 animal growth. Among the 238 DEGs of the meta-analysis, the number of mRNA that were 401 two folds higher in older than younger muscle were 42, 17 and 201 for groups 1, 2 and 3. Of 402 these and when a maximum of 30 mRNA were considered by group, ENO1 and MATR3 were 403 identified in groups 1 and 2; ASPH, STXBP1, RPS12, CYCS, MYOG in groups 2 and 3; and PFKFB3, 404 ATRAID, SDHA, ARPC5, FMO5, PHB, ELN, STRA6, TIMP1 in groups 1 and 3 (data not shown). 405 The number of genes down regulated more than 2 folds were 10, 3 and 10 for groups 1, 2 and 406 3. Of these, TMED9 was identified in groups 2 and 3 and CAPN1, ENO1, MATR3 in groups 1 407 and 3 (data not shown).

408

409 **Discussion**

410 Post-natal muscle growth requires the concerted regulation of numerous genes to ensure the 411 hypertrophy, but also their metabolic and contractile differentiations. Moreover, muscle is 412 not just composed of muscular contractile cells. Satellite cells, adipocytes, fibrocytes, 413 endothelial cells, pericytes, non-satellite stem cells within the muscle... also contribute to 414 muscle growth by hypertrophy (increase in cell volume) and hyperplasia (increase in cell 415 number; (10)). The rates and waves of growth for the different muscular cell types vary 416 according to the age and at similar age according to the breed (28). Thus, the proportion of 417 muscular, fibrous and adipose (also called marbling) cells determines the muscle composition 418 with consequences on meat-eating qualities and thus on the economic value of bovine meat. 419 Using publicly available microarray data from muscles of both high muscling (Belmont Red, 420 Piedmontese x Hereford), high marbling (Angus, Wagyu x Hereford), and intermediate 421 (Chinese Red Step) bovine breeds, we carried out a meta-analysis to identify DEGs between 2 422 post-natal ages disturbed from 10 to 62 percent of mature weight, also cited as 0.1 to 0.62 423 DoM. As the main results, we identified 238 DEG between two ages whatever the bovine 424 breeds. Of these, 97 were identified as DEG in at least 2 independent studies and in the meta-425 analysis, when we carried out a one-to-one dataset analysis, while 17 were never described 426 as DEG between 2 ages. Among the DEG, we emphasized on mRNA that could be 427 transcriptional signatures of muscular lipid deposition as the result of regulated metabolisms 428 or of the differentiation of news adipocytes.

429

430

431 Transcriptional signature of the dynamic regulation of glycolytic and oxidative metabolisms.

432 Besides protein synthesis, glycolytic and oxidative metabolisms are the pathways that sustain 433 the postnatal muscle growth. Muscles use both carbohydrates and fatty acids as energy 434 sources, according to a ratio that differs between glycolytic and oxidative fibers of a muscle, 435 and within a muscle, according to the age. An increased oxidative metabolism is generally 436 concomitant to an increase in marbling (9, 12). Accordingly, we detected a muscular 2-fold 437 mean overexpression for genes related to glycolysis (GPI, PDHA1, PGAM2, PGK1, PKM) and 438 tricarboxylic acid cycle (IDH3B, MDH2, OGDH, SDHA, SUCLG1) when bovines from 0.32 and 439 0.18 DoM were compared (group 1), and for the 5 breeds. However, the abundances of these 440 10 mRNA were similar between bovines at 0.52 or 0.10 DoM (group 2), while they were 2.5-441 (glycolytic) to 5.5-(oxidative) fold higher in muscle from bovine at 0.62 than 0.22 DoM (group 442 3). These results may reflect known variation in the metabolic properties of muscle regarding 443 puberty that generally occurs in cattle at a DoM of approximately 0.5. The DoM near puberty 444 represents a period when an inflection point was reported in the quadratic and cubic 445 relationships of ICDH (tricarboxylic acid cycle enzyme) and LDH (glycolytic enzyme) activities 446 regressed with the DoM of the bovines in several breeds (61). The dynamic nature of the 447 abundance of mRNA related to glycolytic and oxidative metabolisms with DoM may reflect 448 the use of glycolytic metabolism to supply energy during periods of rapid growth before 449 puberty, and the use of glycolytic and more importantly oxidative metabolisms during muscle 450 ageing probably in relation to an increased proportion of lipid deposit (10). We observed a 451 similar dynamic nature of the mRNA abundance related to fast glycolytic or slow oxidative 452 fibers. Indeed MYH1 mRNA (coding for MyHC-2x isoform), related to fast glycolytic fiber was 453 overexpressed in groups 1 and 3 while the slow-type mRNAs such as TNNT1 related to oxidative fibers was overexpressed in group 3. 454

From the current meta-analysis, we proposed that the dynamic regulation of the glycolytic and the oxidative metabolisms may result from the signaling of two muscular master genes MYOG and MAPK14. The myogenin (MYOG) and the Mitogen-activated protein kinase p38 alpha (MAPK14) mRNA were among the top-fold change according to the age, and with a high regulatory potential suggested by high centralities within the network. MYOG and MAPK14 460 are key transcription factors of muscle-specific genes, and they are recognized markers of differentiated fibers and of differentiating satellite cells (29, 47, 57, 64). However, an 461 462 increased MYOG and MAPK14 mRNA abundance during muscle growth and after puberty (top 463 fold change in group3 with an increase of 9 and 10-fold) is much less documented than an 464 overexpression during myogenesis, which raises the question of the significance of MYOG and 465 MAK14 persistent expression in adult bovine muscle deciphered by the current study. One 466 possible explanation for MYOG overexpression in bovine muscle especially after puberty, is to 467 shift the muscle enzymatic activity from glycolytic towards oxidative metabolism, as it was 468 reported in several mice models (26).

469 In differentiated contractile muscular fibers, MAPK14 was shown to be a key regulatory gene 470 of glucose and insulin signaling pathways in muscle from monogastric species (69, 72). 471 Additionally, TXNIP may contribute or sign the enhanced oxidative metabolism. TXNIP was 472 found as an mRNA with a high fold-change and as a gene close to chromosomal region with a 473 QTL of meat yield. The gene TXNIP, encodes the thioredoxin-interacting protein that has been 474 shown to regulate both insulin-dependent and insulin-independent pathways of glucose 475 uptake in human skeletal muscle (48). The current meta-analysis quantified a TXNIP mRNA 476 over-expression of 2.6 and 5.6 fold in the groups 1 and 3 of DoM, which parallels the 477 overexpression of glycolytic mRNA however with a higher fold-change. Thus, MOYG, MAPK14 478 and TXNIP may contribute to the dynamic regulation of glycolytic and oxidative metabolisms 479 within muscle of bovine from 0.1 to 0.62 of DoM.

480

481 Transcriptional signature relating to lipid metabolism

482 Of the 17 DEGs between 2 ages that were identified exclusively by the meta-analysis, APOE, 483 LDLR, MXRA8 and HSP90AA1 upregulation may be related to lipid metabolism and enhanced 484 marbling. Indeed, within muscle, circulating triacylglycerol as VLDL and chylomicrons 485 contribute to ATP production within the contractile muscular cells and to energy storage 486 within adipocytes in bovine through lipoprotein lipase (LPL) activity (11). APOE and LDLR are 487 among the regulatory proteins involved in the lipids delivery to muscle. Apolipoprotein E 488 (APOE) is synthesized primarily by the liver and the adipose tissue but also by the muscle, and 489 it regulates blood lipid and lipoprotein levels in multiple ways. For example, APOE acts as a 490 ligand for low-density lipoprotein receptor (LDLR) and affects the activities of LPL (36).

491 Furthermore, accumulated data report that APOE contributes to the adipogenesis by inducing 492 the assembly of triglycerides in the adipocytes (43). LDLR mRNA abundance was shown to be 493 high in heart but also present in skeletal muscle and adipose tissues according to an age-494 related pattern in mice (71). Studies using LDLR-deficient and transgenic mice have provided 495 evidence for the contribution of LDLR in VLDL-triglyceride metabolism in peripheral tissue 496 including muscle, and for triglyceride storage in the adipocytes (67). Interestingly, another 497 major physiological regulator of triacylglycerol metabolism through LPL activity in muscle is 498 the Angiopoietin-like protein 4 (ANGPTL4) protein. The current study reported a 2- (APOE and 499 LDLR) or 4.5- (ANGPTL4) fold overexpression of mRNA abundance in 0.62 compared to 0.22 500 DoM bovines (group 3). These overexpressions may sign a regulation of triacylglycerol 501 breakdown and storage within the muscular adipose cells; if we assumed that in bovine as in 502 monogastric species these mRNA are much more abundant in adipose than muscular cells. In agreement with this hypothesis of transcriptional signature of adipose cells, LIPE and FABP4 503 504 mRNA abundances followed the same pattern as APOE and LDLR mRNA and were among the 505 238 DEG identified by the current meta-analysis. LIPE is an adipose enzyme involved in the 506 lipolysis of triacylglycerol stored within adipocytes, and a balance between triacylglycerol 507 synthesis and lipolysis was repeatedly observed during lipid deposition (68). COPG1 is a 508 coatomer GTPases involved in the lipolysis signaling within skeletal muscle (19) and was also up-regulated with the age in the present study. FABP4, is a well-known adipose cell marker 509 510 within the muscle, and its mRNA abundance was shown as up-regulated with increased 511 marbling of bovine muscle (33, 39).

512

513 MXRA8 and HSP90AA1 mRNA abundance were also identified among the up-regulated 514 DEG between two ages (2 and 2.5 fold for MXRA8 and HSP90AA1 between bovines of 0.62 515 and 0.22 or 0.52 and 0.1 DoM). MXRA8 is a transmembrane protein that can modulate activity 516 of various signalling pathways (30). However, to date a link between MXRA8 and lipid 517 metabolism is only documented by one publication. MXRA8 was indeed shown as an 518 upregulated gene when bovines with high carcass adiposity and marbling were compared with 519 slightly adipose bovines (18). The Heat shock protein HSP 90-alpha (HSP90AA1), was related 520 to muscular physiology by regulating the transforming growth factor β (TGF- β), by folding the 521 myosin motor domain and by organizing the filament structure of myofibers (24). Inhibition

22

522 of HSP90AA1 function was reported to block TGF_β-induced signaling and transcriptional 523 responses through an enhanced ubiquitination and degradation of Smad3 in various cell lines 524 (65). In bovine, a reduced myostatin abundance, a member of the TGFβ family was associated 525 to a higher muscle mass and to a lower intramuscular fat content (1). Thus, the up-regulation 526 of HSP90AA1 (one of the mRNA with the highest fold change and centralities) in the muscle of 527 ageing bovine, may contribute to the up-regulation of TGF^β with a putative positive impact on 528 adipogenesis or lipid metabolism. In agreement with these various signaling related to an 529 increased lipids deposition with age, we reporte on overexpression of TNFBI mRNA in 0.62 530 compared to 0.22 DoM bovines. In bovine muscle, a decrease abundance of TNFβI mRNA was 531 associated to an increased longissimus muscle area and a decreased marbling (14).

532

533 Transcriptional signature of a balance between adipogenic and myogenic lineages

534 Muscle grows by hypertrophy of muscular contractile cell and by hyperplasia and hypertrophy 535 of other myogenic cell types among them the most important being satellite cells. Indeed, the 536 number of fibers being fixed at birth in bovine (50), satellite cells are a well-known muscle 537 resident myogenic progenitor recruited for muscle growth. Once recruited, satellite cells 538 expressed muscle-specific transcriptional factors including MYOG and MAPK14 found 539 upregulated in groups 2 and mainly 3 for MYOG and in group 3 for MAPK14. However, an age-540 related decrease in satellite cell activity is thought in bovine as in monogastric species, and 541 may partly result from a MAPK14 upregulation. Indeed, inhibition of MAPK14 signaling was 542 shown to maintain bovine satellite cells stemness (22) and to decrease the adipogenic trans-543 differentiation of C2C12 myoblasts (52). Beside satellite cells, the fibrocyte/adipocyte 544 progenitors are a well-known muscle resident cell population recruited for muscle growth 545 under the control of several proteins, of these the bone morphogenetic protein 7 (BMP 7) (10, 546 62). The Y-box-binding protein (YBX1) depletion was shown to lead to a loss of Bmp7 547 expression and a decreased adipogenesis in mice (49). The protein 14-3-3 eta encoded by the 548 gene YWHAH is involved in a large spectrum of both general and specialized signaling 549 pathways. siRNA-mediated silencing of YWHAH led to an increased number of adipose cells 550 differentiated from mesenchymal skeletal stem cells (21). The gene ERBB2 encodes the 551 receptor tyrosine-protein kinase erbB-2, that was shown to regulate the expression of 552 acetylcholine receptors, glucose transport within normal skeletal muscle, as well as to favor 553 the activation of quiescent satellite cells in mice (27) and protein synthesis in culture bovine satellite cells treated with testosterone analog (70). Moreover, the ErbB2 signaling 554 555 downregulation was proposed to enhance the differentiation of adipogenics progenitors 556 through an upregulation of Akt pathway and a downregulation of Erk-1 (63). MAPK14, YBX1, 557 YWAH, ERBB2 were among the mRNA with the highest fold change and with the highest 558 centralities and are thus genes with high regulatory potentials through various pathways. The 559 (50), satellite cells are a well-known muscle resident myogenic progenitor recruited for muscle 560 growth. Strong up regulation of those genes especially in muscle from 0.62 compared to 0.22 561 DoM bovines, may favour (YBX1) or repress (MAPK14, YWAH, ERBB2), depending on the 562 presence of their downstream targets, the differentiation of myogenic progenitors into 563 adipogenic ligneage. Collectively, these results highlight a probable fine-tuning of gene 564 expressions for a muscle growth probably involving several cell populations and a cross-talk 565 between cell types.

566

567 Conclusion

The present study was designed to identify DEGs between two different ages in growing 568 569 bovine muscle, using meta-analysis of multiple microarray datasets throughout the animal 570 lifespan in five breeds or crossbreeds. The difference between two ages was used as a proxy 571 of intramuscular lipid content variation, in the absence of enough muscular microarray 572 datasets carried out in bovines divergent only for marbling. In total, we identified 238 DEGs 573 between two ages; of these, numerous are related to the known dynamic regulation of 574 contractile and metabolic properties of muscle. Network and fold change ranking analyses 575 highlighted some potential "hub" genes that were proposed to contribute to the regulation of 576 dynamic contractile and metabolic pathways. We also highlighted transcriptional signatures 577 of lipid metabolism and of the contribution of muscle-resident progenitor cells to the 578 myogenic or the adipogenic growth with possible impact on the lean-to-fat ratio within muscle 579 and thus on marbling. These findings have practical implications as they pave the way through 580 the second step of the marbling biomarkers discovery pipeline, that is the qualification step 581 (56). Indeed, we provide a list of DEGs valid for different breeds that are regulated in muscle 582 when bovine after and before puberty are compared as a proxy of marbling. The qualification

583 step requires the demonstration that the differential abundance of a candidate biomarker 584 observed in the current meta-analysis driven discovery phase is seen using an alternative 585 method. In the case of mRNA, qPCR should be a valuable method to confirm current DEGs as 586 differentially abundant between groups of bovine divergent for marbling and also to 587 investigate the relationship between the mRNA abundance and the intramuscular fat content. 588 Thus, the present meta-analysis achieved the expected goals: the verification and validation 589 of independent results and the production of conclusive findings related to transcriptional 590 signatures of lipid deposition. Numerous genes were not discussed in this paper and are 591 delivered to meat physiologists interested in the regulation of protein deposition or at the 592 regulation of growth of the connective tissue.

593

594 Acknowledgements

- 595 We thank Isabelle Cassar-Maleck and Brigitte Picard for helpful discussion.
- 596

597 Grants

598 We thank Region Auvergne-Rhône-Alpes and INRAe-Phase Division for funding J.Bazile grant.

599

600 Disclosures

601 The authors declare no conflicts of interest, financial or otherwise.

602 Authors contributions

JB and MB conceived and designed research; FJ and DL mentored JB for performing metaanalysis and statistical analysis. PD and CK performed gene name update. MR helped on raw public data reticule use. JB retrieve public data and performed meta-analysis. JB and MB analyzed data, drafted and wrote the manuscript. FJ, PD, CK and DL edited, revised and approved final version of manuscript

608

Bibliography

- 1. Aiello D, Patel K, Lasagna E. The myostatin gene: an overview of mechanisms of action and its relevance to livestock animals. *Anim Genet* 49: 505–519, 2018.
- 2. Albert M, Andler JA, Bah T, Barbry-Blot P, Barraud J-F, Barton C, Baxter B, Beard J, Bintz J, Biro A, Bishop N, Blocher JL, Böck H, Boczkowski T, Bohre H, Boldewyn, Borgmann D, Bouclet B, Breuer H, Broberg G, Brown C, Brubaker M, Bruno L, Brynn, Buculei N, Byak B, Caclin P, Caldwell I, Carmichael G, Catmur E, Celorio C, Cenoz JA, Ceuppens J, Chyla Z, Clausen A, Cliff J, Cook K, Cromwell B, Crosbie R, Cruz J, De-Cooman A, Derezynski M, Díaz D, Dilly B, Doolittle L, Dufour N, Dwyer T, Dziumanenko M V., Eberl M, Engelen J, Erdelyi M, Erikson U, Falzon N, Faubel S, Felfe F, Fitzsimon A, Flick E, Floryan M, Fowler B, Fred, Gemy C, Giannini S, Gondouin O, Gould T, de Greef T, Grosberg M, De Gussem K, Harrington B, Harvey D, Adnauer Heckert A, Hetherington C, Hirth J, Hochreiner H, Holder T, Holdsworth J, Holmstedt C, Horkan A, Ove Hufthammer K, Hughes R, Hurst N, Inductiveload, Ingham T, Irisson J-O, Jamison B, Janeczko T, Jeanmougin M, Jurema FLB, Kaplinski L, Kerby L, Kiirala N, Kilfiger J, Kitaev N, Kivlighn J, Knoth A, Kosiński K, Kovar P, Lavorata B, Leone A, Leray J, Levien R, van Lierop D, Lindgren N, Lipatov V, Louette I, Marc P-A, Marmion A-A, Marquardt C, Marshall C, Masár I, Mastrukov DG, Mathog D, Matiphas, McDermott P, Meeks M, Mena F, MenTaLguY, Monnier A, Montagne V, Mooney T, Moore DP, Morgan C, Moulder P, Müller J, Nakai Y, Navez V, Neumair C, Nick, Nilsson A, Oka M, dos Santos Oliveira V, Owens M, Penner A, Petroff M, Phillips J, Podobny Z, Prokoudine A, Reinhard J-R, Remizov A, Rodrigo F, Rodrigues H, Rudsatz J, Conde Rueda X, Corrêa da Silva F, Schaller C, Scholten M, von Schwerdtner T, Šegan D, Sharma A, Shivaken, Sloan M, Smith J, Špetič B, Spike A, Sridharan K, Stephan R, Stojek D, Sucha M, Suwalski P, Taraben A, Tebby H, Termeau J, Turner D, Twupack A, Urošević A, Valavanis A, Verona J, Vieites L, Wagenaar D, White LP, Wüst S, Wybrow M, Xg G, Yacob D, Yamato M, Yip D. Inkscape [Online]. Free Software Foundation: 2007. https://www.inkscape.org.
- 3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 215: 403–410, 1990.
- Aranda B, Blankenburg H, Kerrien S, Brinkman FSL, Ceol A, Chautard E, Dana JM, De Las Rivas J, Dumousseau M, Galeota E, Gaulton A, Goll J, Hancock REW, Isserlin R, Jimenez RC, Kerssemakers J, Khadake J, Lynn DJ, Michaut M, O'Kelly G, Ono K, Orchard S, Prieto C, Razick S, Rigina O, Salwinski L, Simonovic M, Velankar S, Winter A, Wu G, Bader GD, Cesareni G, Donaldson IM, Eisenberg D, Kleywegt GJ, Overington J, Ricard-Blum S, Tyers M, Albrecht M, Hermjakob H. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat Methods* 8: 528–529, 2011.

- Baik M, Kang HJ, Park SJ, Na SW, Piao M, Kim SY, Fassah DM, Moon YS. Triennial growth and development symposium: Molecular mechanisms related to bovine intramuscular fat deposition in the longissimus muscle. J Anim Sci 95: 2284–2303, 2017.
- Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics 15: 293, 2014.
- 7. Baron D, Dubois E, Bihouée A, Teusan R, Steenman M, Jourdon P, Magot A, Péréon Y, Veitia R, Savagner F, Ramstein G, Houlgatte R. Meta-analysis of muscle transcriptome data using the MADMuscle database reveals biologically relevant gene patterns. *BMC Genomics* 12: 113, 2011.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res* 41: 991–995, 2013.
- 9. **Bazile J, Picard B, Chambon C, Valais A, Bonnet M**. Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gel-based and gel-free proteomic analyses. *Meat Sci* 156: 146–155, 2019.
- 10. Bonnet M, Cassar-Malek I, Chilliard Y, Picard B. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* 4: 1093–1109, 2010.
- 11. Bonnet M, Faulconnier Y, Hocquette J-F, Bocquier F, Leroux C, Martin P, Chilliard Y. Nutritional status induces divergent variations of GLUT4 protein content, but not lipoprotein lipase activity, between adipose tissues and muscles in adult cattle. *Br J Nutr* 92: 617–625, 2004.
- Bonnet M, Faulconnier Y, Leroux C, Jurie C, Cassar-Malek I, Bauchart D, Boulesteix P, Pethick D, Hocquette JF,
 Chilliard Y. Glucose-6-phosphate dehydrogenase and leptin are related to marbling differences among Limousin and
 Angus or Japanese Black x Angus steers. J Anim Sci 85: 2882–2894, 2007.
- 13. **Casel P, Moreews F, Lagarrigue S, Klopp C**. sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters. *BMC Proc* 3: S3, 2009.
- 14. **Casey TM**, **Walker JF**, **Bhide K**, **Thimmapuram J**, **Schoonmaker JP**. Global transcriptional differences in myokine and inflammatory genes in muscle of mature steer progeny are related to maternal lactation diet and muscle composition. *Physiol Genomics* 50: 884–892, 2018.
- 15. **Ceciliani F, Lecchi C, Bazile J, Bonnet M**. Proteomics Research in the Adipose Tissue. In: *Proteomics in Domestic Animals: from Farm to Systems Biology*. Springer International Publishing, p. 233–254.

- 16. Chaze T, Meunier B, Chambon C, Jurie C, Picard B. In vivo proteome dynamics during early bovine myogenesis. *Proteomics* 8: 4236–4248, 2008.
- 17. **Chaze T, Meunier B, Chambon C, Jurie C, Picard B**. Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle. *Animal* 3: 980–1000, 2009.
- Clark DL, Boler DD, Kutzler LW, Jones KA, McKeith FK, Killefer J, Carr TR, Dilger AC. Muscle Gene Expression Associated with Increased Marbling in Beef Cattle. *Anim Biotechnol* 22: 51–63, 2011.
- Covington JD, Galgani JE, Moro C, LaGrange JM, Zhang Z, Rustan AC, Ravussin E, Bajpeyi S. Skeletal muscle perilipin
 3 and coatomer proteins are increased following exercise and are associated with fat oxidation. *PLoS One* 9: 1–8, 2014.
- 20. Dayanidhi S, Lieber RL. Skeletal muscle satellite cells: Mediators of muscle growth during development and implications for developmental disorders. *Muscle and Nerve* 50: 723–732, 2014.
- 21. **De S, Kline D**. Evidence for the requirement of 14-3-3eta (YWHAH) in meiotic spindle assembly during mouse oocyte maturation. *BMC Dev Biol* 13: 1, 2013.
- 22. Ding S, Swennen GNM, Messmer T, Gagliardi M, Molin DGM, Li C, Zhou G, Post MJ. Maintaining bovine satellite cells stemness through p38 pathway. *Sci Rep* 8: 1–12, 2018.
- 23. Dodson M V. Evolution of meat animal growth research during the past 50 years. (2015). doi: 10.2527/jas2014-8221.
- 24. **Etard C, Armant O, Roostalu U, Gourain V, Ferg M, Strähle U**. Loss of function of myosin chaperones triggers Hsf1mediated transcriptional response in skeletal muscle cells. *Genome Biol* 16: 1–20, 2015.
- 25. **Farhadian M, Rafat SA, Hasanpur K, Ebrahimie E**. Transcriptome signature of the lactation process, identified by meta-analysis of microarray and RNA-Seq data. *BioTechnologia* 99: 153–163, 2018.
- 26. **Flynn JM, Meadows E, Fiorotto M, Klein WH**. Myogenin regulates exercise capacity and skeletal muscle metabolism in the adult mouse. *PLoS One* 5, 2010.
- 27. **Golding JP, Calderbank E, Partridge TA, Beauchamp JR**. Skeletal muscle stem cells express anti-apoptotic ErbB receptors during activation from quiescence. 13, 2006.
- Gotoh T, Albrecht E, Teuscher F, Kawabata K, Sakashita K, Iwamoto H, Wegner J. Differences in muscle and fat accretion in Japanese Black and European cattle. *Meat Sci* 82: 300–308, 2009.
- 29. **Guo L, Cui H, Zhao G, Liu R, Li Q, Zheng M, Guo Y, Wen J**. Intramuscular preadipocytes impede differentiation and promote lipid deposition of muscle satellite cells in chickens. *BMC Genomics* 19: 1–14, 2018.

- 30. Han SW, Jung YK, Lee EJ, Park HR, Kim GW, Jeong JH, Han MS, Choi JY. DICAM inhibits angiogenesis via suppression of AKT and p38 MAP kinase signalling. *Cardiovasc Res* 98: 73–82, 2013.
- 31. **He H, Liu X**. Characterization of Transcriptional Complexity during Longissimus Muscle Development in Bovines Using High-Throughput Sequencing. *PLoS One* 8: e64356, 2013.
- 32. **Jeong J, Bong J, Kim GD, Joo ST, Lee HJ, Baik M**. Transcriptome changes favoring intramuscular fat deposition in the longissimus muscle following castration of bulls. *J Anim Sci* 91: 4692–4704, 2013.
- Jurie C, Picard B, Hocquette JF, Dransfield E, Micol D, Listrat A. Muscle and meat quality characteristics of Holstein and Salers cull cows. *Meat Sci* 77: 459–466, 2007.
- 34. Kaspric N, Picard B, Reichstadt M, Tournayre J, Bonnet M. ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. *PLoS One* 10: e0128086, 2015.
- 35. Keady SM, Kenny DA, Ohlendieck K, Doyle S, Keane MG, Waters SM. Proteomic profiling of bovine M. longissimus lumborum from Crossbred Aberdeen Angus and Belgian Blue sired steers varying in genetic merit for carcass weight. J Anim Sci 91: 654–665, 2013.
- 36. Kersten S. Physiological regulation of lipoprotein lipase. *Biochim Biophys Acta Mol Cell Biol Lipids* 1841: 919–933, 2014.
- 37. Khan DR, Landry DA, Fournier É, Vigneault C, Blondin P, Sirard M-A. Transcriptome meta-analysis of three follicular compartments and its correlation with ovarian follicle maturity and oocyte developmental competence in cows. *Physiol Genomics* 48: 633–643, 2016.
- 38. Kolde R. Pheatmap: Pretty Heatmaps. R package version 1.0.12. .
- 39. Komolka K, Ponsuksili S, Albrecht E, Kühn C, Wimmers K, Maak S. Genomics Data Gene expression pro fi le of Musculus longissimus dorsi in bulls of a Charolais × Holstein F 2 -cross with divergent intramuscular fat content. GDATA 7: 131–133, 2016.
- 40. Lee S, Park E, Cho Y, Kim S, Lee J, Jeon J, Lee C, Im S, Oh S, Thompson JM, Yoon D. Identification of Differentially Expressed Genes Related to Intramuscular Fat Development in the Early and Late Fattening Stages of Hanwoo Steers. 40: 757–764, 2007.
- 41. Lehnert SA, Reverter A, Byrne KA, Wang Y, Nattrass GS, Hudson NJ, Greenwood PL. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev Biol* 7: 95, 2007.

- 42. **Lehnert SAA, Wang YHA, Byrne KAA**. Development and application of a bovine cDNA microarray for expression profiling of muscle and adipose tissue.
- 43. Li YH, Liu L. Apolipoprotein e synthesized by adipocyte and apolipoprotein e carried on lipoproteins modulate adipocyte triglyceride content. *Lipids Health Dis* 13: 1–7, 2014.
- 44. Listrat A, Lebret B, Louveau I, Astruc T, Bonnet M, Lefaucheur L, Bugeon J. How muscle structure and composition determine meat quality. *Prod Anim* 28, 2016.
- 45. **Marot G, Foulley J-L, Mayer C-D, Jaffrezic F**. Moderated effect size and P-value combinations for microarray metaanalyses. *Bioinformatics* 25: 2692–2699, 2009.
- 46. Moisá SJ, Shike DW, Graugnard DE, Rodriguez-Zas SL, Everts RE, Lewin HA, Faulkner DB, Berger LL, Loor JJ. Bioinformatics analysis of transcriptome dynamics during growth in Angus cattle longissimus muscle. *Bioinform Biol Insights* 7: 253–270, 2013.
- 47. Pallafacchina G, Blaauw B, Schiaffino S. Role of satellite cells in muscle growth and maintenance of muscle mass.
 Nutr Metab Cardiovasc Dis 23: 1–7, 2013.
- Parikh H, Carlsson E, Chutkow WA, Johansson LE, Storgaard H, Poulsen P, Saxena R, Ladd C, Schulze PC, Mazzini
 MJ, Jensen CB, Krook A, Björnholm M, Tornqvist H, Zierath JR, Ridderstråle M, Altshuler D, Lee RT, Vaag A, Groop
 LC, Mootha VK. TXNIP regulates peripheral glucose metabolism in humans. *PLoS Med* 4: 0868–0879, 2007.
- 49. **Park JH, Kang HJ, Kang SI, Lee JE, Hur J, Ge K, Mueller E, Li H, Lee BC, Lee SB**. A Multifunctional protein, EWS, is essential for early brown fat lineage determination. *Dev Cell* 26: 393–404, 2013.
- Picard B, Lefaucheur L, Berri C, Duclos MJ. Muscle fibre ontogenesis in farm animal species. *Reprod Nutr Dev* 42: 415–431, 2002.
- 51. **Purslow PP, Archile-Contreras AC, Cha MC.** Meat science and muscle biology symposium: Manipulating meat tenderness by increasing the turnover of intramuscular connective tissue. *J Anim Sci* 90: 950–959, 2012.
- 52. Qi R, Liu H, Wang Q, Wang J, Yang F, Long D, Huang J. Expressions and Regulatory Effects of P38/ERK/JNK Mapks in the Adipogenic Trans-Differentiation of C2C12 Myoblasts. *Cell Physiol Biochem* 44: 2467–2475, 2018.
- 53. Qin L, Zhang G, Cao Y, Zhang J, Zhao Y, Zhao Z. Microarray Analysis on the Differences of Gene Expression in Longissimus Dorsi Muscle Tissue Between 1 and 24 Months Chinese Red Steppes. *Anim Vet Adv* 10: 428–436, 2011.
- 54. **R Core Team**. R: A Language and Environment for Statistical Computing [Online]. R Foundation for Statistical Computing: 2018. https://www.r-project.org/ [9 Apr. 2019].
- 55. **Ramasamy A, Mondry A, Holmes CC, Altman DG**. Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. *PLoS Med* 5: e184, 2008.
- 56. **Rifai N, Gillette MA, Carr SA**. Protein biomarker discovery and validation : the long and uncertain path to clinical utility. 24: 971–983, 2006.
- 57. **Roberts MD**, Haun CT, Mobley CB, Mumford PW, Romero MA, Roberson PA, Vann CG, McCarthy JJ. Physiological differences between low versus high skeletal muscle hypertrophic responders to resistance exercise training: Current perspectives and future research directions. *Front Physiol* 9: 1–17, 2018.
- 58. **Romao JM**, **Jin W**, **He M**, **Mcallister T**, **Guan LL**. Elucidation of Molecular Mechanisms of Physiological Variations between Bovine Subcutaneous and Visceral Fat Depots under Different Nutritional Regimes. 8: 1–14, 2013.
- 59. Rung J, Brazma A. Reuse of public genome-wide gene expression data. Nat Rev Genet 14: 1–11, 2013.
- 60. Sadkowski T, Jank M, Motyl T, Agricultural W. Age-dependent changes in bovine skeletal muscle transcriptomic profile 1.
- 61. Schreurs NM, Garcia F, Jurie C, Agabriel J, Micol D, Bauchart D, Listrat A, Picard B. Meta-analysis of the effect of animal maturity on muscle characteristics in different muscles, breeds, and sexes of cattle. *J Anim Sci* 86: 2872–2887, 2008.
- Schulz TJ, Huang TL, Tran TT, Zhang H, Townsend KL, Shadrach JL, Cerletti M, McDougall LE, Giorgadze N, Tchkonia
 T, Schrier D, Falb D, Kirkland JL, Wagers AJ, Tseng YH. Identification of inducible brown adipocyte progenitors residing in skeletal muscle and white fat. *Proc Natl Acad Sci U S A* 108: 143–148, 2011.
- 63. Scioli MG, Bielli A, Gentile P, Mazzaglia D, Cervelli V, Orlandi A. The Biomolecular Basis of Adipogenic Differentiation of Adipose-Derived Stem Cells. 1: 6517–6526, 2014.
- 64. **Segalés J, Perdiguero E, Muñoz-Cánoves P**. Regulation of muscle stem cell functions: A focus on the p38 MAPK signaling pathway. *Front Cell Dev Biol* 4: 1–15, 2016.
- 65. **Shang Y, Xu X, Duan X, Guo J, Wang Y, Ren F, He D, Chang Z**. Hsp70 and Hsp90 oppositely regulate TGF-β signaling through CHIP/Stub1. *Biochem Biophys Res Commun* 446: 387–392, 2014.
- 66. **Song Y**. Growth Factors and Cytokines in Skeletal Muscle Development, Growth, Regeneration and Disease. Springer International Publishing.
- 67. **Tacken PJ, Hofker MH, Havekes LM, Van Dijk KW**. Living up to a name: The role of the VLDL receptor in lipid metabolism. *Curr Opin Lipidol* 12: 275–279, 2001.

- 68. **Taga H, Bonnet M, Picard B, Zingaretti MC, Cassar-Malek I, Cinti S, Chilliard Y**. Adipocyte metabolism and cellularity are related to differences in adipose tissue maturity between holstein and charolais or Blond d'aquitaine fetuses. *J Anim Sci* 89: 711–721, 2011.
- 69. Tang X, Deng L, Xiong H, Li G, Lin J, Liu S, Xie J, Liu J, Kong F, Tu G, Peng H, Liang S. Expression profile of mitrogenactivated protein kinase (MAPK) signaling genes in the skeletal muscle & liver of rat with type 2 diabetes: role in disease pathology. [Online]. *Indian J Med Res* 140: 744–55, 2014. http://www.ncbi.nlm.nih.gov/pubmed/25758573.
- 70. **Thornton KJ**, **White ME**, **Dayton WR**. Active G protein coupled receptors (GPCR), matrix metalloproteinases epidermal growth factor receptor (EGFR), erbB2, and insulin-like growth factor 1 receptor (IGF-1R) are necessary for trenbolone acetate induced alterations in protein turnover. 9: 2332–2343, 2016.
- 71. **Tiebel O, Oka K, Robinson K, Sullivan M, Martinez J, Nakamuta M, Ishimura-Oka K, Chan L**. Mouse very low-density lipoprotein receptor (VLDLR): Gene structure, tissue-specific expression and dietary and developmental regulation. *Atherosclerosis* 145: 239–251, 1999.
- 72. **Tremblay F, Dubois MJ, Marette A.** Regulation of GLUT4 traffic and function by insulin and contraction in skeletal muscle. *Front Biosci* 8, 2003.
- 73. **Tseng GC, Ghosh D, Feingold E**. Comprehensive literature review and statistical considerations for microarray metaanalysis. *Nucleic Acids Res* 40: 3785–3799, 2012.
- 74. Vaudel M, Verheggen K, Csordas A, Raeder H, Berven FS, Martens L, Vizcaíno JA, Barsnes H. Exploring the potential of public proteomics data. *Proteomics* 16: 214–225, 2016.
- Wang YH, Bower NI, Reverter A, Tan SH, De Jager N, Wang R, McWilliam SM, Cafe LM, Greenwood PL, Lehnert SA.
 Gene expression patterns during intramuscular fat development in cattle. J Anim Sci 87: 119–130, 2009.
- Wang YH, Bower NI, Reverter A, Tan SH, De Jager N, Wang R, McWilliam SM, Cafe LM, Greenwood PL, Lehnert SA.
 Gene expression patterns during intramuscular fat development in cattle. J Anim Sci 87: 119–130, 2009.
- 77. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Sel Work Terry Speed* 30: 591–600, 2012.

Figures captions

- Figure 1 : Prisma flowchart of the data selection. In total, 10 studies comparing muscular transcript abundance between 2 ages of bovine, were included in the meta-analysis. The characteristics of the included studies were given in Table 1. N is the number of datasets/publications selected.
- Figure 2 : Groups of physiological age. Each study is colored according to the group it belongs to: blue for group1, orange for group2 and green for group3. Study E5 is in dot line as the weight of the animals was not specified.E1, E7 and E10 are superposed. The first group is composed by studies E1, E2, E3 and E4 that compared muscles from bovine from 0.18 up to 0.36 of adult weight or DoM. The second group is composed by studies E5, E6, E7, E8 and E9 that compared bovines from 0.1 to 0.52 of DoM. The third group with the study E10 that reported results from bovine muscles from 0.22 to 0.62 of DoM.
- Figure 3 : Clusters of 238 DEGs from the meta-analysis. The clustering was made using the k-means method and heatmap was produced accordingly to this clustering using R package pheatmap. Gene ontology enriched terms identification. The results obtained with ProteINSIDE and enrichments were expressed as –log10 (p-value) to visually plot them on graphs, which means that a -log10 (P value) of 3, 2, and 1.3 correspond to a p-value of 0.001, 0.01, and 0.05, respectively. The genes or the number of genes within the GO annotation were written next to the bar.
- Figure 4: Mean profiles of log ratio of the gene expression by cluster. The number in abscissa corresponds tothe number of the study (1=E1, 2=E2,..,10=E10).15
- Figure 5: Venn diagram (6) highlighting genes among the 238 DEG of the meta-analysis that were found with
the highest betweeness centrality (BC), Closeness centrality (CC) and Fold change (FC).18

DISCUSSION, CONCLUSIONS ET PERSPECTIVES

9 DISCUSSION GENERALE

L'objectif de cette thèse était **d'agréger et de réanalyser des données publiques pour proposer des gènes et protéines marqueurs de l'adiposité corporelle ou musculaire, compléter les connaissances disponibles ou identifier des manques de données**. Nous avons répondu à ces objectifs en confirmant ou en identifiant pour la première fois des protéines et ARNm musculaires dont les variations d'abondances sont concomitantes de variation de l'adiposité corporelle ou musculaire (LIM). Outre l'apport de connaissance, une seconde originalité de ce travail de thèse est la combinaison d'approches *in silico* et expérimentale et aussi la combinaison de méthodes de protéomiques.

9.1 LES DONNEES PROTEOMIQUES POUR LA COMPREHENSION DE LA COMPOSITION CORPORELLE ET MUSCULAIRE DES BOVINS

Une approche *in silico* a permis, par la synthèse des résultats de 5 publications, la production d'une liste de 50 protéines dont l'abondance varie avec la teneur en LIM. Parmi ces 50 protéines, 9 ont été retrouvées dans au moins deux publications. Cependant ces publications portaient essentiellement sur des bovins de races précoces pour les dépôts adipeux. Aussi, nous avons voulu compléter les données disponibles sur le changement de protéome concomitant avec une variation de l'adiposité, mais cette fois sur des races tardives. Afin de maximiser nos chances de découverte de protéines et aussi pour compléter la description du protéome musculaire, nous avons utilisé deux techniques de protéomique, la 2DE et le shotgun, ce qui a rarement été fait, et jamais chez le bovin.

Nous avons ainsi mis en évidence 47 protéines dont l'abondance dans le muscle corrèle avec la teneur en LIM ou avec le pourcentage de tissu adipeux dans la carcasse. Parmi ces protéines, 21 étaient communes à celles identifiées chez les races précoces et 26 ont été identifiées pour la première fois. Les protéines communes signent des régulations, des métabolismes et des modifications contractiles des fibres musculaires, avec une diminution des voies glycolytiques et une augmentation des voies oxydatives en race Rouge des Prés. Parmi les résultats originaux obtenus au cours de la thèse, nous avons montré une relation positive entre l'abondance d'ALDH1A1 et APOBEC2 et la teneur en LIM, ainsi qu'avec l'adiposité de la carcasse pour ALDH1A1. ALDH1A1 est une protéine impliquée dans le métabolisme de la vitamine A, qui pourrait contribuer à l'orientation du métabolisme oxydatif en augmentant l'abondance de l'alpha-B cristalline (CRYAB) aussi quantifiée dans notre étude. APOBEC2 est une protéine spécifique du muscle, trouvée préférentiellement dans les muscles lents oxydatifs, et a été montrée comme liée à la variation de LIM chez les bovins dans notre étude et en parallèle chez des races précoces (Thornton et al., 2017).

Le différentiel d'abondance de ALDH1A1 et de CRYAB selon la teneur en LIM a été confirmé par des méthodes immunologique et de spectrométrie de masse ciblée. Les abondances ainsi quantifiées ont montré que l'abondance de ALDH1A1 ou de CRYAB contribuerait respectivement à 18 ou 11 % de la variabilité en LIM dans le *Longissimus thoracis* de bovins Rouge des Prés (Bonnet et al., soumis).

Simultanément à la thèse et après la soumission de l'article 1, de nouvelles publications ont rapporté des différences d'abondance de protéines en lien avec une variation des LIM. Thornton, Chapalamadugu, Eldredge et Murdoch (2017) (2-DE), ont décrit la différence de protéome du *Longissimus thoracis* de bovins divergents par la note de conformation (grille USDA) et le persillé. La race n'est pas précisée, et est probablement une race précoce, car les carcasses correspondent au marché américain. De Souza Rodrigues et al., (2017), ont décrit (2-DE) la différence de protéome entre des bovins Angus (Bos Taurus) et Nellore (Bos Indicus, des zébus).

Discussion, conclusions et perspectives



Figure 29: Diagramme de Venn des protéines dont l'abondance varie avec la teneur en LIM. Produit en utilisant <u>http://jvenn.toulouse.inra.fr/app/example.html</u> (Bardou, Mariette, Escudié, Djemiel et Klopp, 2014). Les protéines utilisées de (Bazile, Picard, Chambon, Valais et Bonnet, 2019) sont uniquement celles corrélées au dépôt adipeux intramusculaire. Pour rappel, l'étude Ceciliani est la synthèse des 5 publications de protéomiques disponible jusqu'en 2017.

Lorsque la totalité des protéines différentiellement abondantes décrites en octobre 2019 sont comparées (Figure 29), TPI1, une enzyme de la glycolyse, est la seule protéine identifiée dans toutes les études. En outre, l'abondance de TPI1 diminue lorsque la teneur en LIM augmente quelle que soit la race ou les conduites d'élevage des bovins. La plus faible abondance de TPI1 dans des muscles avec de plus fortes teneurs en LIM, a été confirmée par des méthodes immunologique et de spectrométrie de masse ciblée, et elle contribuerait à 18 % de la variabilité en LIM en race Rouge des Près (Bonnet et al., soumis). En accord avec ce résultat, MYL1, la chaîne légère de la Myosine 1, caractéristique des fibres glycolytiques a été identifiée comme sous-abondante lorsque la teneur en lipides augmente et ceci dans au moins 5 études. Enfin, lorsque l'on considère les protéines identifiées dans au moins 3 études, 7 autres protéines sont listées avec cependant, un sens de variation d'abondance qui diffère entre les études (Tableau 16). Ces données confirment que les protéines les plus souvent décrites avec des abondances différentes selon la teneur en LIM sont des protéines liées aux métabolismes ou aux propriétés contractiles du muscle. Enfin, il est probable que les données produites par les méthodes plus sensibles de protéomiques que la 2-DE, telles que la méthode de shotgun, révèlent des protéines moins classiquement identifiées, potentiellement régulatrices des propriétés contractiles et métaboliques du muscle, comme illustré par nos résultats avec ALDH1A1.

Tableau 16: Parmi les protéines comparées dans la figure X, 9 ont été identifiées dans au moins 3 publications. Le sens de variation de l'abondance des protéines est reporté en fonction de l'augmentation de la teneur en lipides intramusculaires. Les 3 protéines surlignées en vert sont parmi les 9 identifiées dans au moins 2 des 5 études utilisées dans Ceciliani et al. (2017).

| ∕ de LIM | (Ceciliani, Lecchi, Bazile et Bonnet, 2018) | (De Souza Rodrigues et al., 2017) | (Thornton et al., 2017) | (Bazile et al., 2019) |
|----------|---|---|-------------------------|--------------------------|
| TPI1 | Ч | R | И | Ц |
| MYL1 | Ч | Ы | И | |
| PGM1 | Ч | лv N | $\not \rightarrow$ | |
| ENO1 | | Ы | И | Л |
| ACTA1 | | И | 7 | Z |
| TNNT1 | Z | | И | Л |
| PGK1 | Я | | Ц | Ы |
| SDHA | Ы | | 7 | Л |
| MYH1 | Ы | | 7 | Ы |

9.2 LA REUTILISATION DES DONNEES PUBLIQUES VIA LA META-ANALYSE

La réutilisation de données publiques transcriptomiques a présenté plusieurs défis. La gestion des milliers de données liées à l'expression de gènes (jusqu'à 24 000 gènes différents)

a incité à la mise en place d'outils d'archivage et de versionning des données et des analyses. L'utilisation d'outils de versionning tel que Git permet un suivi des analyses ainsi que des fichiers et surtout de leurs modifications. En outre, dans les bases de données transcriptomiques, aucune des métadonnées liées aux jeux de données ne rapportaient les valeurs liées à l'adiposité corporelle ou au pourcentage de LIM. Aussi, j'ai recherché ces informations dans les publications et/ou auprès des auteurs directement, en demandant en particulier les teneurs en LIM et le poids de carcasse chaude. En intégrant le poids de carcasse chaude aux équations de prédiction développées par Al-Jammas, (2017), j'ai estimé l'adiposité d'une carcasse. A l'issue de ce travail, 9 jeux de données sur 43 avaient des données d'adiposité corporelle ou musculaire. Aussi, les critères d'adiposité n'ont pas pu être utilisés pour catégoriser les données, et nous avons choisi de les catégoriser selon les facteurs de variation majeurs de l'adiposité : l'âge, la race, le sexe et l'alimentation. Les métadonnées relatives aux jeux de données ont été compilées dans un tableur en vue de la description et de la publication des données.

Le format de certaines données déposés dans la base de données transcriptomique GEO a été particulièrement contraignant. En effet, pour certains jeux de données nous n'avions pas accès aux données brutes, mais uniquement aux données archivées sous le format de log ratio. Ce type de format concernait en particulier le GSE5659, qui apportait environ la moitié des données et les écarter aurait affaibli la méta-analyse. Nous avons donc fait le choix de les inclure et de calculer les log ratio entre deux conditions pour toutes les données. Le format log ratio des abondances d'ARNm a orienté le choix de la méthode de méta-analyse vers la combinaison des valeurs p. Nous avons alors initié une collaboration avec Florence JAFFREZIC et Denis LALOE (UMR GABI, Inra, Jouy-en-Josas) qui m'ont formée à l'utilisation d'un paquet R dédié à la combinaison des valeurs p en prenant comme matrice d'entrée les données en log ratio. Nous avons pu adapter un script R d'analyse de ces données et mettre en place une méthodologie d'analyse.

Le format des identifiants de gènes (ID) était aussi très différent d'un jeu de données à l'autre probablement en raison de la diversité des puces ADN utilisées ou de l'obtention des données sur plus de 10 ans. Le génome bovin ayant évolué sur cette période, nous avons fait le choix de mettre à jour tous les ID en comparant la séquence des sondes ADN à la dernière version (UCD1.2) du génome bovin rendue disponible début 2018. Pour cette mise à jour, nous avons bénéficié de l'aide et des outils de Patrice DEHAIS et Christophe KLOPP (Sigenae, Inra, Toulouse). J'ai géré ce projet comme un projet à part entière, référencé sous l'acronyme Borderer. Tous les ID n'ont pas pu être mis à jour, car certains auteurs ne donnaient pas accès

aux séquences des sondes ADN. Pour les ID mis à jour, environ 40% du nombre de gènes par étude a été conservé. Cette réduction du nombre de gènes a réduit le nombre de gènes potentiellement communs entre les études à inclure dans la méta-analyse.

Après agrégation et préparation des données catégorisées selon les 4 facteurs de variation de l'adiposité, nous avons analysé les abondances d'ARNm dans le muscle de bovin selon le facteur de variation majeur qui est l'âge des animaux. Sur les 715 gènes inclus dans la méta-analyse car communs à 10 jeux de données, 238 gènes étaient différentiellement exprimés selon l'âge. Parmi ceux-ci 17 gènes ont été montrés comme différentiellement exprimés uniquement dans la méta-analyse, et n'étaient pas déclarés différentiels quand les études étaient analysées une à une. Par contre, 97 gènes étaient identifiés comme différentiellement exprimés dans au moins 2 études et la méta-analyse. Parmi les DEG en fonction de l'âge, nous avons identifié des ARNm qui pourraient signer des variations de la teneur en LIM, du fait de leur rôle dans les métabolismes lipidiques en particulier dans le métabolisme des triacylglycerols circulants (APOE, LDLR, MXRA8, FAPB4, LIPE) tandis que d'autres pourraient induire (YBX1) ou réprimer (MAPK14, YWAH, ERBB2) la différenciation de cellules progénitrices musculaires vers le lignage adipeux. Ces résultats sont tout à fait originaux et n'avaient pour la plupart pas été mis en évidence dans les 5 études publiées rapportant les transcriptomes différentiels selon l'âge des bovins (Moisá et al. (2013) Qin et al. (2011), Wang et al., (2009), Sadkowski et al., (2014) et Komolka et al., (2016). En outre, lorsque la liste des 238 DEG identifiés dans la méta-analyse a été comparée à celles issues des publications de Wang et al., (2009), Sadkowski et al., (2014) et Komolka et al., (2016) seuls 3 gènes ont été retrouvés dans la méta-analyse et au moins une autre étude (Erreur ! Source du renvoi introuvable.) : FNTB (pourrait être impliqué via la signalisation de la rhodopsine dans la régulation du métabolisme oxydatif), MYL3 (un gène qui code pour une protéine analogue à MYL1) et FABP4 (un marqueur des cellules adipeuses dans le muscle très souvent corrélés à la teneurs en lipides

Discussion, conclusions et perspectives

(Jurie et al., 2007)).



Figure 30: Diagramme de Venn des gènes différentiellement exprimés et communs entre les études rapportant la différence de transcriptome selon l'âge des bovins. Les résultats de Moisá et al. (2013) et de Qin et al. (2011) n'ont pas pu être inclus dans la comparaison en raison respectivement de l'absence de liste de gènes DEG, et d'ID obsolète. Le faible consensus que nous observons entre les études provient probablement d'une différence d'ID associés à chaque publication, et du faible nombre d'ARNm listé dans les publications relativement aux données déposées dans GEO.

160

De même, 23 gènes ont été retrouvés en commun entre Wang et al., (2009) et la métaanalyse : ATF4, PFKFB3, TXNIP, GSTM3, ADRB2, LAMA4, MGP, FOS, CSRP3, LDHA, FN1, ATP2A1, APOE, CA3, ACTN3, MYH1, CLU, COX7A2, CKM, FHL1, TPM2, ACTB et H3F3A. Comme attendu et en accord avec la synthèse réalisée par Baik et al (2017), intégrant aussi des résultats de PCR, ces gènes signent des régulations des propriétés contractiles du muscle (ACT3, ACTB, TPM2) ou de la transcription des gènes (ATF4, FOS, FN1, CSRP3), selon l'âge des bovins.

9.3 Les donnees omiques combinees, un atout pour comprendre L'Adiposite Musculaire ?

La comparaison des données transcriptomiques et protéomiques présente deux intérêts majeurs :

- Un intérêt biologique, puisqu'il indique si une variation de l'abondance de l'ARNm impacte l'abondance de la protéine qui est « l'unité fonctionnelle » de la cellule capable de modifier un caractère biologique.
- Un intérêt plus appliqué dans une perspective de qualification des viandes, avec l'opportunité de cibler soit l'ARNm soit la protéine selon les résultats de la relation entre l'abondance de l'ARNm ou de la protéine et la teneur en LIM.

Nous avons donc comparé la totalité des données protéomiques et les 238 DEG selon l'âge identifié par la méta-analyse (Figure 31).



Figure 31 : Diagramme de Venn des protéines différentiellement abondantes selon la teneur en lipides et des gènes différentiellement exprimés selon l'âge identifiés par la méta-analyse. Produit en utilisant <u>http://jvenn.toulouse.inra.fr/app/example.html (Bardou et al., 2014)</u> et modifié en ajoutant les noms des gènes/protéines communes aux données transcriptomiques et protéomiques. Les protéines utilisées de (Bazile et al., 2019)(nommées Bazile_protéines) sont uniquement celles corrélées au dépôt adipeux intramusculaire.

Comme illustré sur la Figure 31, (238-212=)23 gènes dont l'expression musculaire varie avec l'âge codent pour des protéines dont l'abondance varie selon la teneur en lipides (certaines différences de LIM ayant été induites par l'âge). Parmi ceux-ci, 5 sont des protéines classiquement identifiées comme différentielles selon la teneur en LIM du muscle : ENO1,

TNNT1, PGK1, SDHA et MYH1 (Tableau 16). Sans surprise, ces gènes signent des modifications, selon l'âge ou la teneur en LIM, de l'expression de gènes (ARNm et protéines) qui codent des protéines liées :

- A la glycolyse et au métabolisme du glucose : PKM, GPI, ENO1, PGAM2, PGK1,
- Aux propriétés contractiles du muscle ACTN3, TNNT1, TPM1, MYBPC1, TPM2 et MYL3, et MYH1
- Au cycle de Krebs avec SDHA et OGDH, ou au métabolisme oxydatif CYCS
- Et la myoglobine, le gène le plus souvent différentiellement exprimé dans la métaanalyse, les études transcriptomiques (Lehnert et al., 2007; H. B. Wang, Zan et Zhang, 2014) et les études protéomique (Bazile et al., 2019 et Thornton et al., 2017)

La combinaison des résultats de transcriptomiques et protéomiques ciblent ces 23 gènes comme des gènes potentiellement liés à des modifications de la composition du muscle. Ils suggèrent d'analyser la relation entre l'abondance de ces 23 gènes et la teneur en lipides du muscle, afin de sélectionner un ensemble d'ARNm susceptibles d'expliquer ou de prédire la variabilité de la teneur en LIM du *Longissimus thoracis*. Enfin, si des relations mathématiques existent entre les abondances d'ARNm et la teneur en LIM, la quantification multiplex d'ARNm pourrait être une solution de phénotypage moléculaire des LIM. Un tel outil pourrait être largement déployé car l'abondance d'ARNm est plus simple et moins couteuse à mesurer que des abondances de protéines.

10 CONCLUSIONS ET PERSPECTIVES

Grace à la réutilisation de données publiques transcriptomiques et protéomiques, complétées par des analyses protéomiques, ce travail de thèse a confirmé le lien entre les propriétés métaboliques et contractiles du muscle et la teneur en LIM. Il a aussi identifié pour la première fois dans un contexte d'adiposité musculaire, des gènes dont l'expression modulerait la teneur en LIM du fait de régulations de métabolismes des lipides ou du nombre de cellules adipeuses dans le muscle. Ce travail de thèse a ciblé 47 protéines et 238 DEG potentiellement marqueurs de la teneur en LIM, dont 23 gènes/protéines communes, ce qui ouvre des perspectives de recherche et d'application. La première perspective est d'étudier la

relation entre l'abondance des 23 ARNm et la teneur en LIM du muscle *Longissimus thoracis* de nombreux bovins d'âges et de races différentes, selon le protocole de recherche de biomarqueurs proposé par Rifai, Gillette et Carr, (2006). Pour cette étude l'utilisation de PCR multiplex permettrait de tester une méthode haut-débit et sensible compatible avec la nécessité de vérifier les relations abondances-LIM sur de larges cohortes de bovins. Une deuxième perspective est de croiser la liste des 23 gènes avec celles des protéines musculaires (Picard et Gagaoua, soumis) et plasmatiques (Boudon et al., soumis) liées à la tendreté de la viande dans de nombreuses races. Ceci permettrait de comprendre les déterminants biologiques communs à la tendreté et à la teneur en LIM, et aussi de cibler des indicateurs pour qualifier simultanément deux critères majeurs des qualités de viande bovine.

Ce travail de thèse a largement utilisé des données publiques. La réutilisation des données publiques est un véritable défi qui s'inscrit dans la démarche mondiale de réduction d'utilisation des animaux lors d'expérimentation selon le principe des 3R. Elle s'inscrit aussi dans un effort international de réduction de la redondance des recherches et d'une meilleure exploitation des données existantes. Une telle démarche outre la synthèse, la vérification ou la production de connaissances, identifie aussi les manques de connaissances et oriente les recherches à réaliser. La démarche d'analyse systématique de la bibliographie requiert des méthodes et compétences diverses. La réalisation de ce travail de thèse a aussi produit une méthodologie, avec des scripts d'analyses de résultats (Annexe) et de visualisation des données utiles aux chercheurs. Elle a aussi agrégé, mis à jour et organisé de nombreuses données stockées dans GEO, données que nous rendrons accessibles aux physiologistes du muscle bovin sous la forme de data paper. Ainsi, ces données pourront être analysées avec d'autres objectifs de recherche que ceux de la compréhension des déterminants de l'adiposité musculaire.

Annexes

ANNEXES



Auvergne-Rhône-Alpe

Meta-analysis of public transcriptomics data to understand and phenotype bovine body composition



Bench test of the R script



Université Clermont Auvergne, INRA, VetAgro Sup, UMR Herbivores, F-63122 Saint-Genès-Champanelle, France ² INRA, UMR1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France

Whole body and muscular adiposities contribute to the animal adaptability, food efficiency and meat/carcass qualities. To date, the challenge is to integrate the **numerous genomics studies** that have been published in order to identify key pathways and biomarkers of adiposities whatever the rearing practices.

Our aim is to aggregate public data and to implement statistical tool in order to identify robust biomarkers of the adipose tissue deposition.

Material et methods

Public Transcriptomics dataset









A mong twelve thousand identifiers for each studies, only 855 were common and included in the meta-analysis

In collaboration with Sigenae (www.sigenae.org/, INRA), we currently re-annotate the datasets on the actual bovine to increase the number of genes included in the comparison

HOXA5 (to be confirmed by the ongoing annotation) is an embryonic factor reported to be expressed in adipose tissues of adult rat but has never been related to post-natal muscle biology. We proposed HOXA5 as a potential driver or marker of muscle accretion whatever

Take home message

We provide a pipeline of aggregation of public transcriptomics data to foster the discovery of key genes in order to understand or phenotype bovine body composition



POINTS CLÉS DU SCRIPT D'ANALYSE DES RÉSULTATS DE PROTEOMIQUE : MANAGEMENT OF PROTEOMIC DATA TO IDENTIFY DIFFERENTIALLY ABUNDANT PROTEINS ACCORDING TO ONE DISCRIMINANT FACTOR AND TO CORRELATE THEIR ABUNDANCE WITH THE VALUE OF THIS FACTOR

1 CONTEXTE

- Existence de nombreux logiciels ou script d'analyses des données protéomiques : MaxQuant (Perseus), Skyline ou Progenesis ou script R tel que MSstats, DAPAR, ProStar, SafeQuant.
- Mais outils sous licence et/ou dont les méthodes statistiques ne sont pas toujours bien caractérisées.
- Nécessité d'une méthode d'analyse pour identifier les protéines différentiellement abondantes entre deux conditions, et pour analyser la relation entre l'abondance d'une protéine et une valeur quantitative du facteur de variation.

2 OBJECTIFS

Création d'un script et mise à disposition de la communauté scientifique et nonscientifique pour :

- Pour accepter ou rejeter Ho de l'absence de différence entre 2 groupes
 - (ici l'adiposité)
- Calculer la corrélation entre l'abondance d'une protéine et une valeur quantitative du facteur de variation

3 DEMARCHE EXPERIMENTALE/ANALYTIQUE

p-values p-values p-values du test de Student pour provenant du provenant du Test de Shapiro test de test de les données rapportant la Student **Kruskal-Wallis** normales ou du distribution normal quelque soit la quelque soit la test de Kruskal-(ou non)de distribution des distribution Wallis test pour les l'abondance (TRUE= des données données données nondistribution normale) (TRUE= pval (TRUE= pval normales (TRUE= significative) significative) pval significative) pval.K W pval.testTor signif.testTorK pval.shapiro pval.Ttest signif.Ttest signif.KW normality кw w Protein1 0.089 TRUE 0.801 FALSE 0.801 FALSE 0.465 FALSE 0.403 TRUE TRUE 0.040 TRUE 0.047 TRUE Protein2 0.040 Protein3 0.018 FALSE 0.047 TRUE 0.061 TRUE 0.047 TRUE Protein4 0.000 FALSE 0.175 FALSE 0.675 FALSE 0.175 FALSE

3.1 IDENTIFICATION DES VALEURS DIFFERENTIELLEMENT ABONDANTES

→ Les résultats sont enregistrés sur l'ordinateur de l'utilisateur au format csv

$10.1\ \text{Correlation}$ avec la valeur discriminante

Création d'une figure comprenant les protéines significativement corrélées avec la valeur de corrélation (positive ou négative) et la valeur de corrélation de pearson ou de Kendall/Spearman (à déterminer par l'utilisateur)





→ Production de 2 tableaux (enregistré au format csv) comprenant

- Les valeurs de corrélation
- Les valeurs p des corrélations

Tableau 17: exemple de tableau avec les valeurs de corrélations pour un jeu test

| | adiposity | Protein1 | Protein2 | Protein3 | Protein4 |
|-----------|-------------|-------------|-------------|-------------|-----------------|
| adiposity | 1 | 0,06723532 | 0,82398759 | -0,8845269 | ۔ 0,15145677 |
| Protein1 | 0,06723532 | 1 | -0,0077549 | -0,04223508 | ۔ 0,04956437 |
| Protein2 | 0,82398759 | -0,0077549 | 1 | -0,60280322 | 0,245925 |
| Protein3 | -0,8845269 | -0,04223508 | -0,60280322 | 1 | 0,16626722 |
| Protein4 | -0,15145677 | -0,04956437 | 0,245925 | 0,16626722 | 1 |

4 CONCLUSIONS

- Méthodes statistiques à choisir selon la distribution normale ou non.
- Production d'un script facile d'utilisation, accessible à tous, disponible à l'adresse :

https://github.com/jane-bzl/Differential_abundance_and_correlation_to_one_factor

| Impliqué dans la comparaison N° | Provenant du nSE | Age (jours) | Sexe | Race | Nombres d'animaux | Poids du corps vivant (kg) | Poids de la carcasse chaude (kg) | Poids de Carcasse théorique de l'adulte (kg) | Age physiologique | % tissue adipeux | % tissue musculaire | persillé/ %LIM | But | Maturité | Nombre de gène minimum (avant mise à jour) | Nombre de gene minimum (après mise à jour) | Pourcentage de perte du nombre de gène (%) |
|------------------------------------|------------------|-------------|------|---------------------------|-------------------|----------------------------|-------------------------------------|---|-------------------|------------------|---------------------|----------------|--------|-------------------|---|---|---|
| 6-8 | 19586 | 1080 | F | Qinchuan | 3 | NS | NS | 420 | NS | NS | NS | NS | Viande | Précoce | 24125 | 15177 | 37 |
| 6-7 | 19587 | 1080 | CM | Qinchuan | 3 | NS | NS | NS | NS | NS | NS | NS | Viande | Précoce | 24125 | 15177 | 37 |
| 7-8 | 19588 | 1080 | В | Qinchuan | 3 | NS | NS | 820 | NS | NS | NS | NS | Viande | Précoce | 24125 | 15177 | 37 |
| 23 | 25554 | 92 | F | Wagyu x Hereford | 4 | 92.1 | 50.7 | 510 | 0.18 | NS | NS | NS | Viande | Précoce | 18785 | 9916 | 47 |
| 23 | 25554 | 93 | F | Piedmontese x Hereford | 4 | 93.6 | 52.4 | 533 | 0.18 | NS | NS | NS | Viande | Intermé diaire | 18785 | 9916 | 47 |
| 25 | 25554 | 366 | F | Wagyu x Hereford | 2 | 267.5 | 147.1 | 510 | 0.52 | NS | NS | NS | Viande | Précoce | 18958 | 9977 | 47 |
| 25 | 25554 | 366 | F | Piedmontese x Hereford | 2 | 269.3 | 150.8 | 533 | 0.51 | NS | NS | NS | Viande | Intermé diaire | 18958 | 9977 | 47 |
| 26 | 25554 | 610 | F | Wagyu x Hereford | 2 | 398.4 | 219.9 | 510 | 0.78 | NS | NS | 7.2 | Viande | Précoce | 18495 | 9803 | 47 |
| 26 | 25554 | 610 | F | Piedmontese x Hereford | 2 | 403 | 226.9 | 533 | 0.76 | NS | NS | 7.3 | Viande | Intermé diaire | 18495 | 9803 | 47 |
| 27 | 25554 | 763 | F | Wagyu x Hereford | 2 | 468.6 | 260.1 | 510 | 0.92 | NS | NS | NS | Viande | Précoce | 18294 | 9749 | 47 |

Tableau 18 : Description des paramètres animaux liés au log ratio (comparaison de 2 conditions) des jeux de données transcriptomiques.

| 27 | 25554 | 763 | F | Piedmontese x Hereford | 2 | 476.6 | 269.8 | 533 | 0.89 | NS | NS | NS | Viande | Intermé diaire | 18294 | 9749 | 47 |
|----|-------|--------------------|----|---------------------------|---|-------|-------|-----------------|------|------|----|------|--------|-------------------|-------|-------|----|
| 28 | 25554 | 915 | F | Wagyu x Hereford | 2 | 677 | 371 | 510 | 1.33 | 23.9 | 56 | 8.77 | Viande | Précoce | 18872 | 9951 | 47 |
| 28 | 25554 | 915 | F | Piedmontese x Hereford | 2 | 658 | 390 | 533 | 1.23 | 22.2 | 58 | 5.09 | Viande | Intermé diaire | 18872 | 9951 | 47 |
| 29 | 25554 | -220 | NS | Wagyu x Hereford | 1 | 0.013 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 18430 | 9781 | 47 |
| 29 | 25554 | -220 | NS | Piedmontese x Hereford | 2 | 0.015 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 18430 | 9781 | 47 |
| 30 | 25554 | -145 | NS | Wagyu x Hereford | 2 | 1.547 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 19044 | 10012 | 47 |
| 30 | 25554 | -145 | NS | Piedmontese x Hereford | 1 | 1.833 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 19044 | 10012 | 47 |
| 31 | 25554 | -85 | NS | Wagyu x Hereford | 2 | 8.587 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 18850 | 9951 | 47 |
| 31 | 25554 | -85 | NS | Piedmontese x Hereford | 1 | 8.107 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 18850 | 9951 | 47 |
| 32 | 25554 | 0 | NS | Wagyu x Hereford | 1 | 31.5 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 18756 | 9886 | 47 |
| 32 | 25554 | 0 | NS | Piedmontese x Hereford | 2 | 36 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 18756 | 9886 | 47 |
| 33 | 25554 | -6 avant régime | NS | Belmont Red | 2 | 211 | NS | B:950, F:550 | NS | NS | NS | NS | Viande | Tardif | 18438 | 9727 | 47 |
| 33 | 25554 | -6 avant régime | NS | Belmont Red | 2 | 210 | NS | B:950, F:550 | NS | NS | NS | NS | Viande | Tardif | 18438 | 9727 | 47 |

| 34 | 25554 | 114 après régime | NS | Belmont Red | 2 | 321 | 161 | B:950, F:550 | NS | 24.3 | 54 | NS | Viande | Tardif | 18841 | 9881 | 48 |
|--------------|-------|------------------------|----|-----------------------|---|-------|-----|-----------------|------|------|----|------|-------------------|---------|-------|-------|----|
| 34 | 25554 | 114 après régime | NS | Belmont Red | 2 | 178 | 78 | B:950, F:550 | NS | NS | NS | NS | Viande | Tardif | 18841 | 9881 | 48 |
| 35 | 25554 | 204 après régime | NS | Belmont Red | 2 | 524.5 | 264 | B:950, F:550 | NS | 18.3 | 60 | 1.85 | Viande | Tardif | 18970 | 9982 | 47 |
| 35 | 25554 | 204 après régime | NS | Belmont Red | 2 | 481 | 243 | B:950, F:550 | NS | 18.8 | 60 | 2.4 | Viande | Tardif | 18970 | 9982 | 47 |
| 36 | 35085 | 610 | S | Korean = Hanwoo | 3 | 671 | NS | B:466, F:357 | 0.9 | NS | NS | 11.3 | Viande | Précoce | 8404 | 5422 | 35 |
| 36 | 35085 | 610 | S | Korean = Hanwoo | 3 | 611 | NS | B:466, F:357 | 0.9 | NS | NS | NS | Viande | Précoce | 8404 | 5422 | 35 |
| 37-38 | 35085 | 846 | В | Korean = Hanwoo | 3 | 671 | 309 | 466 | 0.9 | 16.8 | 62 | 2.6 | Viande | Précoce | 7933 | 5422 | 32 |
| 37-38 | 35085 | 600 | S | Korean = Hanwoo | 3 | 611 | 398 | 466 | 0.9 | 19.8 | 60 | 11.3 | Viande | Précoce | 7933 | 5422 | 32 |
| 39 | 46411 | 458 | В | Hereford | 4 | NS | NS | 820 | NS | NS | NS | 1.1 | Viande | Précoce | 19994 | 11469 | 43 |
| 39-40 | 46411 | 458 | В | Limousine | 4 | NS | NS | 1150 | NS | NS | NS | 0.53 | Viande | Tardif | 19486 | 11268 | 42 |
| 40 | 46411 | 458 | В | Holstein- Friesian | 4 | NS | NS | 1100 | NS | NS | NS | 0.81 | Viande et lait | Précoce | 19486 | 11268 | 42 |
| 41-42- 43 | 48136 | 155 | S | Angus | 7 | 200 | NS | 900 | 0.22 | NS | NS | NS | Viande | Précoce | 13008 | 7306 | 44 |
| 41 | 48136 | 215 | S | Angus | 7 | 265 | NS | 900 | 0.29 | NS | NS | NS | Viande | Précoce | 13128 | 7306 | 44 |
| 42 | 48136 | 275 | S | Angus | 7 | 387 | NS | 900 | 0.43 | NS | NS | NS | Viande | Précoce | 13043 | 7247 | 44 |
| 43 | 48136 | 375 | S | Angus | 7 | 560 | NS | 900 | 0.62 | NS | NS | NS | Viande | Précoce | 13008 | 7245 | 44 |
| 53 | 5659 | -220 | NS | Wagyu x Hereford | 9 | 0.013 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 9371 | 2237 | 76 |

| 57 | 5659 | -220 | NS | Piedmontese x Hereford | 9 | 0.015 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 9371 | 2237 | 76 |
|--------------|------|------|----|---------------------------|---|-------|----|-----|------|----|----|----|--------|-------------------|------|------|----|
| 53-54 | 5659 | -145 | NS | Wagyu x Hereford | 9 | 1.547 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 9371 | 2237 | 76 |
| 57-58 | 5659 | -145 | NS | Piedmontese x Hereford | 9 | 1.833 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 9371 | 2237 | 76 |
| 52-58 | 5659 | -85 | NS | Piedmontese x Hereford | 9 | 8.107 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 9371 | 2237 | 76 |
| 54-59 | 5659 | -85 | NS | Wagyu x Hereford | 9 | 8.587 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 9371 | 2192 | 77 |
| 52 | 5659 | 0 | NS | Piedmontese x Hereford | 9 | 36 | NS | 533 | NS | NS | NS | NS | Viande | Intermé diaire | 9371 | 2237 | 76 |
| 59 | 5659 | 0 | NS | Wagyu x Hereford | 9 | 31.5 | NS | 510 | NS | NS | NS | NS | Viande | Précoce | 9371 | 2192 | 77 |
| 51-60 | 5659 | 92 | F | Wagyu x Hereford | 9 | 92.1 | NS | 510 | 0.18 | NS | NS | NS | Viande | Précoce | 3033 | 995 | 67 |
| 55-60 | 5659 | 92 | F | Piedmontese x Hereford | 9 | 93.6 | NS | 533 | 0.18 | NS | NS | NS | Viande | Intermé diaire | 3033 | 995 | 67 |
| 51-61- 67 | 5659 | 214 | F | Wagyu x Hereford | 9 | 185.9 | NS | 510 | 0.36 | NS | NS | NS | Viande | Précoce | 3033 | 995 | 67 |
| 55-56- 61 | 5659 | 214 | F | Piedmontese x Hereford | 9 | 186.2 | NS | 533 | 0.35 | NS | NS | NS | Viande | Intermé diaire | 3033 | 995 | 67 |
| 65 | 5659 | 244 | S | Belmont Red | 9 | 250 | NS | 950 | 0.26 | NS | NS | NS | Viande | Tardif | 9371 | 2237 | 76 |
| 56-62 | 5659 | 366 | F | Piedmontese x Hereford | 9 | 269.3 | NS | 533 | 0.51 | NS | NS | NS | Viande | Intermé diaire | 3033 | 995 | 67 |
| 62 | 5659 | 366 | F | Wagyu x Hereford | 9 | 267.5 | NS | 510 | 0.52 | NS | NS | NS | Viande | Précoce | 3033 | 995 | 67 |

| 63-64- 65 | 5659 | 366 | S | Belmont Red | 9 | 325 | NS | 950 | 0.34 | NS | NS | NS | Viande | Tardif | 3035 | 999 | 67 |
|--------------|-------|-----|----|------------------------------------|----|-------|-------|------|------|------|----|------|-------------------|---------|-------|-------|----|
| 66 | 5659 | 366 | NS | Belmont Red | 9 | 350 | NS | 950 | 0.37 | NS | NS | NS | Viande | Tardif | 3035 | 995 | 67 |
| 66 | 5659 | 366 | | Belmont Red | 9 | 300 | NS | 950 | 0.32 | NS | NS | NS | Viande | Tardif | 3035 | 995 | 67 |
| 67 | 5659 | 366 | F | Wagyu x Hereford | 9 | 267.5 | NS | 510 | 0.52 | NS | NS | NS | Viande | Précoce | 9371 | 2192 | 77 |
| 63-64 | 5659 | 458 | S | Belmont Red | 9 | 400 | NS | 950 | 0.42 | NS | NS | NS | Viande | Tardif | 3035 | 999 | 67 |
| 69 | 75347 | 549 | В | Charolais x Holstein Friesan | 10 | NS | NS | NS | NS | NS | NS | 7 | Viande | Tardif | 24000 | 15176 | 37 |
| 69 | 75347 | 549 | В | Charolais x Holstein Friesan | 10 | NS | NS | NS | NS | NS | NS | 1.9 | Viande | Tardif | 24000 | 15176 | 37 |
| 73 | 21782 | 732 | В | Chinese Red Steppe | 8 | NS | | 1000 | 0.1 | NS | NS | NS | Viande | Précoce | 23444 | 15177 | 35 |
| 73 | 21782 | 31 | В | Chinese Red Steppe | 8 | NS | | 1000 | 0.4 | NS | NS | NS | Viande et lait | Précoce | 23444 | 15177 | 35 |
| 81 | 52145 | 252 | F | Charolais x Limousin | 7 | 512.7 | 270.2 | 750 | 0.45 | 19.3 | 63 | 3.09 | Viande | Tardif | 23323 | 15177 | 35 |
| 81 | 52145 | 252 | F | Charolais x Limousin | 7 | 506.1 | 276.8 | 750 | 0.44 | 19.2 | 63 | 4.12 | Viande | Tardif | 23323 | 15177 | 35 |
| 83 | 5659 | 329 | S | Holstein | 3 | 246 | NS | 1100 | 0.2 | NS | NS | NS | Viande et lait | Précoce | 3017 | 994 | 67 |
| 83 | 5659 | 363 | S | Japanese Black | 3 | 277 | NS | 809 | 0.3 | NS | NS | NS | Viande | Précoce | 3017 | 994 | 67 |

F=Femelle, B=Taureaux (bulls), CM=mâle castré, NS= non spécifié

174

Annexes

Les races sont catégorisées selon leur précocité à déposer des tissus adipeux ; Précoce ; Intermédiaire et Tardif

Age : exprimé en jours et calculé en multipliant le nombre de mois de l'animal par 30,5. Les valeurs négatives, sont des âges postnataux calculés par soustraction de 280 jours (nombre de jours moyen d'une gestation) au nombre de jour de développement du fœtus.

Annexes

BIBLIOGRAPHIE

- Agreste. (2019). Animaux de boucherie. Bovins cheptel décembre 2018 Resultats français et
européens.européens.Paris.Repéréà
http://agreste.agriculture.gouv.fr/IMG/pdf/2019_034_InforapBovinsv2.pdf
- AgriMer. (2010). Guide du classificateur carcasse.
- Al-Jammas, M. (2017). Prédiction de la composition de la carcasse basée sur le métabolisme des nutriments absorbés : vers une évolution des recommandations alimentaires des jeunes bovins. Université Clermont Auvergne.
- Al-Shahrour, F., Díaz-Uriarte, R. et Dopazo, J. (2004). FatiGO: A web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 20(4), 578-580. doi:10.1093/bioinformatics/btg455
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. et Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. doi:10.1016/S0022-2836(05)80360-2
- Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F. S. L., Ceol, A., Chautard, E., ... Hermjakob, H. (2011). PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nature Methods*, 8(7), 528-529. doi:10.1038/nmeth.1637
- Arloth, J., Bader, D. M., Röh, S. et Altmann, A. (2015). Re-Annotator: Annotation pipeline for microarray probe sequences. *PLoS ONE*, 10(10), 1-13. doi:10.1371/journal.pone.0139516
- Aschwanden, C. (2015). Not Even Scientists Can Easily Explain P-values. *Scientific Method*. Repéré à https://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/
- Bader, G. D. (2006). Pathguide: a Pathway Resource List. Nucleic Acids Research, 34(90001), D504-D506. doi:10.1093/nar/gkj126
- Baik, M., Kang, H. J., Park, S. J., Na, S. W., Piao, M., Kim, S. Y., ... Moon, Y. S. (2017). Triennial growth and development symposium: Molecular mechanisms related to bovine intramuscular fat deposition in the longissimus muscle. *Journal of Animal Science*, 95(5), 2284-2303. doi:10.2527/jas2016.1160
- Baldwin, R. L., Li, R. W., Li, C. -j., Thomson, J. M. et Bequette, B. J. (2012). Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. *Physiological Genomics*, 44(10), 543-550. doi:10.1152/physiolgenomics.00144.2011
- Barbosa-Morais, N. L., Dunning, M. J., Samarajiwa, S. A., Darot, J. F. J., Ritchie, M. E., Lynch, A. G. et Tavaré, S. (2009). A re-annotation pipeline for Illumina BeadArrays: Improving the interpretation of gene expression data. *Nucleic Acids Research*, 38(3). doi:10.1093/nar/gkp942
- Bardou, P., Mariette, J., Escudié, F., Djemiel, C. et Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, 15(1), 293. doi:10.1186/1471-2105-15-293
- Baron, D., Dubois, E., Bihouée, A., Teusan, R., Steenman, M., Jourdon, P., ... Houlgatte, R. (2011). Meta-analysis of muscle transcriptome data using the MADMuscle database reveals biologically relevant gene patterns. *BMC Genomics*, *12*(1), 113. doi:10.1186/1471-2164-12-113
- Bauchart, D., Chantelot, F. et Gandemer, G. (2008). Qualités nutritionnelles de la viande et des abats chez le bovin : données récentes sur les principaux constituants d'intérêt nutritionnel. *Cahiers de Nutrition et de Diététique*, 43. Repéré à https://prodinra.inra.fr/record/29734
- Bazile, J., Picard, B., Chambon, C., Valais, A. et Bonnet, M. (2019). Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gel-based and gel-free proteomic analyses. *Meat science*, 156(May), 146-155. doi:10.1016/j.meatsci.2019.05.018

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G. et Brunak, S. (2004). Feature-based prediction

of non-classical and leaderless protein secretion. *Protein Engineering, Design and Selection, 17*(4), 349-356. doi:10.1093/protein/gzh037

- Bhandary, M. et Zhang, X. (2011). Comparison of several tests for combining several independent tests. *Journal of Modern Applied Statistical Methods*, 10(2), 436-446. doi:10.22237/jmasm/1320120240
- Bonnet, M., Cassar-Malek, I., Chilliard, Y. et Picard, B. (2010). Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal*, 4(7), 1093-1109. doi:10.1017/S1751731110000601
- Bonnet, M., Faulconnier, Y., Leroux, C., Jurie, C., Cassar-Malek, I., Bauchart, D., ... Chilliard, Y. (2007). Glucose-6-phosphate dehydrogenase and leptin are related to marbling differences among Limousin and Angus or Japanese Black x Angus steers. *Journal of Animal Science*, 85(11), 2882-2894. doi:10.2527/jas.2007-0062
- Bonnet, Muriel, Faulconnier, Y., Hocquette, J.-F., Bocquier, F., Leroux, C., Martin, P. et Chilliard, Y. (2004). Nutritional status induces divergent variations of GLUT4 protein content, but not lipoprotein lipase activity, between adipose tissues and muscles in adult cattle. *British Journal of Nutrition*, 92(4), 617-625. doi:10.1079/bjn20041240
- Bonny, S. P. F., Hocquette, J. F., Pethick, D. W., Farmer, L. J., Legrand, I., Wierzbicki, J., ... Gardner, G. E. (2016). The variation in the eating quality of beef from different sexes and breed classes cannot be completely explained by carcass measurements. *Animal*, 10(6), 987-995. doi:10.1017/S175173111500292X
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. et Rothstein, H. R. (2010). A basic introduction to fixedeffect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111. doi:10.1002/jrsm.12
- Borenstein, M., Hedges, L. V, Higgins, J. P. T. et Rothstein, H. R. (2009). Introduction to meta-analysis. Leadership and Organizational Outcomes: Meta-Analysis of Empirical Studies. Wiley. doi:10.1007/978-3-319-14908-0_2
- Borstein, S. R. et O'Meara, B. C. (2018). AnnotationBustR: An R package to extract subsequences from GenBank annotations. *PeerJ*, 2018(7). doi:10.7717/peerj.5179
- Buness, A., Huber, W., Steiner, K., Sültmann, H. et Poustka, A. (2005). arrayMagic: Two-colour cDNA microarray quality control and preprocessing. *Bioinformatics*, 21(4), 554-556. doi:10.1093/bioinformatics/bti052
- Buza, T. J., Mccarthy, F. M., Wang, N., Bridges, S. M. et Burgess, S. C. (2008). Gene Ontology annotation quality analysis in model eukaryotes. *Nucleic Acids Research*, *36*(2), 1-9. doi:10.1093/nar/gkm1167
- Caccia, D., Dugo, M., Callari, M. et Bongarzone, I. (2013). Bioinformatics tools for secretome analysis. Biochimica et Biophysica Acta - Proteins and Proteomics, 1834(11), 2442-2453. doi:10.1016/j.bbapap.2013.01.039
- Cafferky, J., Hamill, R. M., Allen, P., O'Doherty, J. V., Cromie, A. et Sweeney, T. (2019). Effect of Breed and Gender on Meat Quality of M. longissimus thoracis et lumborum Muscle from Crossbred Beef Bulls and Steers. *Foods*, 8(5), 173. doi:10.3390/foods8050173
- Caraux, G. et Pinloche, S. (2005). PermutMatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*, 21(7), 1280-1281. doi:10.1093/bioinformatics/bti141
- Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., ... Gaudet, P. (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics*, 25(2), 288-289. doi:10.1093/bioinformatics/btn615

- Carrillo, J. A., He, Y., Li, Y., Liu, J., Erdman, R. A. et Sonstegard, T. S. (2016). Integrated metabolomic and transcriptome analyses reveal finishing forage affects metabolic pathways related to beef quality and animal welfare. *Nature Publishing Group*, (May), 1-16. doi:10.1038/srep25948
- Ceciliani, F., Lecchi, C., Bazile, J. et Bonnet, M. (2018). Proteomics Research in the Adipose Tissue. Dans *Proteomics in Domestic Animals: from Farm to Systems Biology* (p. 233-254). Cham : Springer International Publishing. doi:10.1007/978-3-319-69682-9_12
- Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., ... Coutinho, L. L. (2015). Putative Regulatory Factors Associated with Intramuscular Fat Content. *PLOS ONE*, *10*(6), e0128350. doi:10.1371/journal.pone.0128350
- Cesar, A. S. M., Regitano, L. C. A., Poleti, M. D., Andrade, S. C. S., Tizioto, P. C., Oliveira, P. S. N., ... Coutinho, L. L. (2016). Differences in the skeletal muscle transcriptome profile associated with extreme values of fatty acids content. *BMC Genomics*, 1-16. doi:10.1186/s12864-016-3306-x
- Chaze, T., Meunier, B., Chambon, C., Jurie, C. et Picard, B. (2008). In vivo proteome dynamics during early bovine myogenesis. *Proteomics*, 8(20), 4236-4248. doi:10.1002/pmic.200701101
- Chen, D., Li, W., Du, M., Wu, M. et Cao, B. (2015). Sequencing and Characterization of Divergent Marbling Levels in the Beef Cattle (Longissimus dorsi Muscle) Transcriptome, 28(2), 158-165.
- Chen, J. E. et Glover, G. H. (2016). Panther paper, 25(3), 289-313. doi:10.1007/s11065-015-9294-9.Functional
- Cheng, C., Shen, K., Song, C., Luo, J. et Tseng, G. C. (2009). Ratio adjustment and calibration scheme for gene-wise normalization to enhance microarray inter-study prediction. *Bioinformatics*, 25(13), 1655-1661. doi:10.1093/bioinformatics/btp292
- Chilliard, Y., Delavaud, C. et Bonnet, M. (2005). Leptin expression in ruminants: Nutritional and physiological regulations in relation with energy metabolism. *Domestic Animal Endocrinology*, 29(1), 3-22. doi:10.1016/j.domaniend.2005.02.026
- Choi, S. H., Silvey, D. T., Johnson, B. J., Doumit, M. E., Chung, K. Y., Sawyer, J. E., ... Smith, S. B. (2014). Conjugated linoleic acid (t-10, c-12) reduces fatty acid synthesis de novo, but not expression of genes for lipid metabolism in bovine adipose tissue ex vivo. *Lipids*, 49(1), 15-24. doi:10.1007/s11745-013-3869-0
- Clark, D. L., Boler, D. D., Kutzler, L. W., Jones, K. A., McKeith, F. K., Killefer, J., ... Dilger, A. C. (2011). Muscle Gene Expression Associated with Increased Marbling in Beef Cattle. *Animal Biotechnology*, 22(2), 51-63. doi:10.1080/10495398.2011.552031
- Cochran, W. G. (1954). The Combination of Estimates from Different Experiments. *Biometrics*, *10*(1), 101. doi:10.2307/3001666
- Cundiff, L. V., Szabo, F., Gregory, K. E., Koch, R. M., Dikeman, M. E. et Crouse, J. D. (1993). Breed Comparisons in the Germplasm Evaluation Program At Marc. *Meeting of Beef Improvement Federation*, 17. doi:10.1002/jcph.788.This
- Dawson, D. V., Pihlstrom, B. L. et Blanchette, D. R. (2016). Understanding and evaluating metaanalysis. *Journal of the American Dental Association*, 147(4), 264-270. doi:10.1016/j.adaj.2015.10.023
- De Souza Rodrigues, R. T., Chizzotti, M. L., Vital, C. E., Baracat-Pereira, M. C., Barros, E., Busato, K. C., ... Da Silva Martins, T. (2017). Differences in beef quality between Angus (Bos taurus taurus) and Nellore (Bos taurus indicus) cattle through a proteomic and phosphoproteomic approach. *PLoS ONE*, *12*(1), 1-21. doi:10.1371/journal.pone.0170294
- Debray, T. P. A., Moons, K. G. M., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H. H. et Reitsma, J. B. (2015). Get real in individual participant data (IPD) meta-analysis: A review

of the methodology. Research Synthesis Methods, 6(4), 293-309. doi:10.1002/jrsm.1160

- DeConde, R. P., Hawley, S., Falcon, S., Clegg, N., Knudsen, B. et Etzioni, R. (2006). Combining results of microarray experiments: A rank aggregation approach. *Statistical Applications in Genetics and Molecular Biology*, 5(1). doi:10.2202/1544-6115.1204
- Drennan, M. J., McGee, M. et Keane, M. G. (2008). The value of muscular and skeletal scores in the live animal and carcass classification scores as indicators of carcass composition in cattle. *Animal*, 2(5), 752-760. doi:10.1017/S1751731108001754
- Du, M., Wang, B., Fu, X., Yang, Q. et Zhu, M. J. (2015). Fetal programming in meat production. *Meat Science*, 109, 40-47. doi:10.1016/j.meatsci.2015.04.010
- Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. et Trent, J. M. (1999). Expression profiling using cDNA microarrays. *Nature Genetics*, 21(S1), 10-14. doi:10.1038/4434
- EBI Scientific Report 2017 Contents. (2017). *EMBL-EBI Annual Scientific Report*. Repéré à https://www.embl.de/aboutus/communication_outreach/publications/ebi_ar/ebi_ar_2017.pdf
- Eden, E., Navon, R., Steinfeld, I., Lipson, D. et Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10, 1-7. doi:10.1186/1471-2105-10-48
- ElMasry, G., Kamruzzaman, M., Sun, D. W. et Allen, P. (2012). Principles and Applications of Hyperspectral Imaging in Quality Evaluation of Agro-Food Products: A Review. *Critical Reviews in Food Science and Nutrition*, *52*(11), 999-1023. doi:10.1080/10408398.2010.543495
- Emanuelsson, O., Nielsen, H., Brunak, S. et Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, *300*(4), 1005-1016. doi:10.1006/jmbi.2000.3903
- Emrich, S. J., Barbazuk, W. B., Li, L. et Schnable, P. S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research*, *17*(1), 69-73. doi:10.1101/gr.5145806
- English, A. M., Waters, S. M., Cormican, P., Byrne, C. J., Fair, S. et Kenny, D. A. (2018). Effect of early calf-hood nutrition on the transcriptomic profile of subcutaneous adipose tissue in Holstein-Friesian bulls. *BMC Genomics*, 19(1), 1-12. doi:10.1186/s12864-018-4681-2
- Farhadian, M., Rafat, S. A., Hasanpur, K., Ebrahimi, M. et Ebrahimie, E. (2018). Cross-species metaanalysis of transcriptomic data in combination with supervised machine learning models identifies the common gene signature of lactation process. *Frontiers in Genetics*, 9(JUL), 1-17. doi:10.3389/fgene.2018.00235
- Farhadian, M., Rafat, S. A., Hasanpur, K. et Ebrahimie, E. (2018). Transcriptome signature of the lactation process, identified by meta-analysis of microarray and RNA-Seq data. *BioTechnologia*, 99(2), 153-163. doi:10.5114/bta.2018.75659
- Farmer, L. J. et Farrell, D. T. (2018). Review: Beef-eating quality: A European journey. *Animal*, *12*(11), 2424-2433. doi:10.1017/S1751731118001672
- Feng, W., Wang, G., Sunshine, M., Narasimhan, S., Lababidi, S., Riss, J. et Barrett, C. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biology*.
- Fisher, R. (1925). Statistical methods for research workers. Edinburgh: Genesis Publishing, Oliver and
Boyd;Biological
monographsand
manuals.Repéré
à
à
http://psychclassics.yorku.ca/Fisher/Methods
- Font-I-Furnols, M., Carabús, A., Pomar, C. et Gispert, M. (2014). Estimation of carcass composition and cut composition from computed tomography images of live growing pigs of different genotypes. *Animal*, *17*(1), 166-178. doi:10.1017/S1751731114002237

FRANCEAGRIMER. (2018). Les fiches de FranceAgriMer Filière. Repéré à www.franceagrimer.fr

- Gagaoua, M., Picard, B. et Monteils, V. (2018). Associations among animal, carcass, muscle characteristics, and fresh meat color traits in Charolais cattle. *Meat Science*, *140*(March), 145-156. doi:10.1016/j.meatsci.2018.03.004
- Garbutt, G. J., Anthony, W. B., Walker, D. F. et McGuire, J. A. (1979). Perirectal Adipose Tissue Development of Postweaned Rapidlty Growing Bull Calves. *Journal of Animal Science*, 48(3), 525-530. doi:10.2527/jas1979.483525x
- Gettys, T. W., Henricks, D. M. et Schanbacher, B. D. (1988). An assessment of the relationship between tissue growth patterns and selected hormone profiles among sex phenotypes in cattle. *Animal Production*, 47(3), 335-343. doi:10.1017/S0003356100003469
- Ghosh, D. et Choi, H. (2008). metaArray package for meta-analysis of microarray data Data preparation and POE transformation. *Pnas*, 1-7.
- Giaretta, E., Mordenti, A. L., Canestrari, G., Brogna, N., Palmonari, A. et Formigoni, A. (2018). Assessment of muscle Longissimus thoracis et lumborum marbling by image analysis and relationships between meat quality parameters. *PLoS ONE*, *13*(8), 1-12. doi:10.1371/journal.pone.0202535
- Goldstein, D. R., Delorenzi, M., Luthi-Carter, R. et Sengstag, T. (2011). Comparison of meta-analysis to combined analysis of a replicated microarray study. Dans *Meta-analysis and Combining Information in Genetics and Genomics*.
- Goldstein, H., Yang, M., Omar, R., Turner, R. et Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society*. *Series C: Applied Statistics*, *49*(3), 399-412. doi:10.1111/1467-9876.00200
- Gondret, F. et Hocquette, J. F. (2006). La teneur en lipides de la viande: Une balance métabolique complexe. *Productions Animales*, 19(5), 327-338.
- Gotoh, T., Albrecht, E., Teuscher, F., Kawabata, K., Sakashita, K., Iwamoto, H. et Wegner, J. (2009). Differences in muscle and fat accretion in Japanese Black and European cattle. *Meat Science*, 82(3), 300-308. doi:10.1016/j.meatsci.2009.01.026
- Guerra, R. et Goldstein, D. R. (2016). *Meta-analysis and Combining Information in Genetics and Genomics*. Chapman and Hall/CRC. doi:10.1201/9781420010626
- Gupta, S. K., Bencurova, E., Srivastava, M., Pahlavan, P., Balkenhol, J. et Dandekar, T. (2016). Improving Re-annotation of Annotated Eukaryotic Genomes. Dans *Big Data Analytics in Genomics* (p. 171-195). Hong Kong : Springer International Publishing.
- Gurevitch, J., Koricheva, J., Nakagawa, S. et Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, 555(7695), 175-182. doi:10.1038/nature25753
- Hausman, G. J., Basu, U., Wei, S., Hausman, D. B. et Dodson, M. V. (2014). Preadipocyte and Adipose Tissue Differentiation in Meat Animals: Influence of Species and Anatomical Location. *Annual Review of Animal Biosciences*, 2(1), 323-351. doi:10.1146/annurev-animal-022513-114211
- Hausman, G. J. et Canovas, A. (2017). 2016 triennial growth and development symposium: New perspectives and insight on intramuscular adipose tissue. *Journal of Animal Science*, 95(5), 2241-2243. doi:10.2527/jas2017.1480
- He, H. et Liu, X. (2013). Characterization of Transcriptional Complexity during Longissimus Muscle Development in Bovines Using High-Throughput Sequencing. *PLoS ONE*, 8(6), e64356. doi:10.1371/journal.pone.0064356
- Heard, N. A. et Rubin-Delanchy, P. (2018). Choosing between methods of combining p -values. *Biometrika*, 105(1), 239-246. doi:10.1093/biomet/asx076

- Higgins, J. et Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions. IEEE International Symposium on Information Theory - Proceedings. A John Wiley & Sons, Ltd., Pu.
- Hiller, K., Grote, A., Scheer, M., Münch, R. et Jahn, D. (2004). PrediSi: Prediction of signal peptides and their cleavage positions. *Nucleic Acids Research*, 32(WEB SERVER ISS.), 375-379. doi:10.1093/nar/gkh378
- Hirai, S., Matsumoto, H., Hino, N., Kawachi, H., Matsui, T. et Yano, H. (2007). Myostatin inhibits differentiation of bovine preadipocyte. *Domestic Animal Endocrinology*, 32(1), 1-14. doi:10.1016/j.domaniend.2005.12.001
- Hocquette, J. F., Gondret, F., Baza, E., Mdale, F., Jurie, C. et Pethick, D. W. (2010). Intramuscular fat content in meat-producing animals: Development, genetic and nutritional control, and identification of putative markers. *Animal*, 4(2), 303-319. doi:10.1017/S1751731109991091
- Huang, D. W., Sherman, B. T. et Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44-57. doi:10.1038/nprot.2008.211
- Huntley, R. P., Binns, D., Dimmer, E., Barrell, D., O'Donovan, C. et Apweiler, R. (2009). QuickGO: A user tutorial for the web-based Gene Ontology browser. *Database*, 2009, 1-19. doi:10.1093/database/bap010
- Huntley, R. P., Sawford, T., Martin, M. J. et O'Donovan, C. (2014). Understanding how and why the Gene Ontology and its annotations evolve: The GO within UniProt. *GigaScience*, *3*(1), 1-9. doi:10.1186/2047-217X-3-4
- Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J. et O'Donovan, C. (2015). The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Research*, 43(D1), D1057-D1063. doi:10.1093/nar/gku1113
- Isaksson, T., Nilsen, B. N., Tøgersen, G., Hammond, R. P. et Hildrum, K. I. (1996). On-line, proximate analysis of ground beef directly at a meat grinder outlet. *Meat Science*, 43(3-4), 245-253. doi:10.1016/S0309-1740(96)00016-2
- Jin, W., Olson, E. N., Moore, S. S., Basarab, J. A., Basu, U. et Guan, L. L. (2015). Transcriptome analysis of subcutaneous adipose tissues in beef cattle using 3 ' digital gene expression-tag profiling 1, 171-183. doi:10.2527/jas.2011-4229
- Jurie, C., Picard, B., Hocquette, J. F., Dransfield, E., Micol, D. et Listrat, A. (2007). Muscle and meat quality characteristics of Holstein and Salers cull cows. *Meat Science*, 77(4), 459-466. doi:10.1016/j.meatsci.2007.04.014
- Käll, L., Krogh, A. et Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research*, 35(SUPPL.2), 429-432. doi:10.1093/nar/gkm256
- Karimpour-Fard, A., Epperson, L. E. et Hunter, L. E. (2015). A survey of computational tools for downstream analysis of proteomic and other omic datasets. *Human genomics*, 9, 28. doi:10.1186/s40246-015-0050-2
- Kaspric, N., Picard, B., Reichstadt, M., Tournayre, J. et Bonnet, M. (2015). ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. *PLOS ONE*, 10(5), e0128086. doi:10.1371/journal.pone.0128086
- Keady, S. M., Kenny, D. A., Ohlendieck, K., Doyle, S., Keane, M. G. et Waters, S. M. (2013). Proteomic profiling of bovine M. longissimus lumborum from Crossbred Aberdeen Angus and Belgian Blue sired steers varying in genetic merit for carcass weight. *Journal of Animal Science*, 91(2), 654-665. doi:10.2527/jas.2012-5850

- Keel, B. N., Zarek, C. M., Keele, J. W., Kuehn, L. A., Snelling, W. M., Oliver, W. T., ... Lindholm-Perry, A. K. (2018). RNA-Seq Meta-analysis identifies genes in skeletal muscle associated with gain and intake across a multi-season study of crossbred beef steers. *BMC Genomics*, 19(1), 1-11. doi:10.1186/s12864-018-4769-8
- Kempster, A. J. (1986). Correlations between indirect and direct measurements of body composition. *Proceedings of the Nutrition Society*, 45(1), 55-62. doi:10.1079/pns19860035
- Keogh, K., Kenny, D. A., Cormican, P., Mccabe, M. S., Kelly, A. K. et Waters, S. M. (2016). Effect of Dietary Restriction and Subsequent Re-Alimentation on the Transcriptional Profile of Bovine Skeletal Muscle. doi:10.1371/journal.pone.0149373
- Khan, D. R., Landry, D. A., Fournier, É., Vigneault, C., Blondin, P. et Sirard, M.-A. (2016). Transcriptome meta-analysis of three follicular compartments and its correlation with ovarian follicle maturity and oocyte developmental competence in cows. *Physiological Genomics*, 48(8), 633-643. doi:10.1152/physiolgenomics.00050.2016
- Kim, N. K., Lee, S. H., Cho, Y. M., Son, E. S., Kim, K. Y., Lee, C. S., ... Park, E. W. (2009). Proteome analysis of the m. longissimus dorsi between fattening stages in Hanwoo steer. *BMB Reports*, 42(7), 433-438. doi:10.5483/BMBRep.2009.42.7.433
- Klee, E. W. et Ellis, L. B. M. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics*, 6, 1-7. doi:10.1186/1471-2105-6-256
- Kolde, R., Laur, S., Adler, P. et Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, 28(4), 573-580. doi:10.1093/bioinformatics/btr709
- Komolka, K., Ponsuksili, S., Albrecht, E., Kühn, C., Wimmers, K. et Maak, S. (2016). Genomics Data Gene expression pro fi le of Musculus longissimus dorsi in bulls of a Charolais × Holstein F 2 cross with divergent intramuscular fat content. *GDATA*, 7, 131-133. doi:10.1016/j.gdata.2015.12.023
- La consommation de viande en France. (2018). *planetoscope*. Repéré à https://www.planetoscope.com/elevage-viande/1587-consommation-de-viande-en-france.html
- Lancaster, H. O. (1961). THE COMBINATION OF PROBABILITIES: AN APPLICATION OF ORTHONORMAL FUNCTIONS. *Australian Journal of Statistics*, *3*(1), 20-33. doi:10.1111/j.1467-842X.1961.tb00058.x
- Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., ... Davis, R.
 W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences*, 94(24), 13057-13062. doi:10.1073/pnas.94.24.13057
- Laurioux, B. (2018). La viande, d'hier à aujourd'hui : histoire et enjeux. Dans M.-P. ELLIES-OURY et J.-F. HOCQUETTE (dir.), *La chaîne de la viande bovine. Production, transformation, valorisation et consommation* (Sciences &, p. 1-14). Paris : Lavoisier Tec & Doc.
- Lee, S. C., Lee, H. J., Kim, D. W., Park, J. G. et Han, I. K. (2000). Effects of Carbon Precursors and Hormones on the Lipogenesis and Lipolysis of Hanwoo Cattle Adipose Tissues. *Asian-Australasian Journal of Animal Sciences*, 13(3), 300-306. doi:10.5713/ajas.2000.300
- Lee, Sangdae, Lohumi, S., Lim, H. S., Gotoh, T., Cho, B. K. et Jung, S. (2015). Determination of intramuscular fat content in beef using magnetic resonance imaging. *Journal of the Faculty of Agriculture, Kyushu University*, 60(1), 157-162.
- Lee, Seung-hwan, Gondro, C., Werf, J. Van Der, Kim, N., Lim, D., Park, E., ... Thompson, J. M. (2010). Use of a bovine genome array to identify new biological pathways for beef marbling in Hanwoo (Korean Cattle). *BMC Genomics*, *11*(1), 623. doi:10.1186/1471-2164-11-623

- Leek, J. T., Monsen, E., Dabney, A. R. et Storey, J. D. (2006). EDGE: Extraction and analysis of differential gene expression. *Bioinformatics*, 22(4), 507-508. doi:10.1093/bioinformatics/btk005
- Lehnert, S. A., Reverter, A., Byrne, K. A., Wang, Y., Nattrass, G. S., Hudson, N. J. et Greenwood, P. L. (2007). Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Developmental Biology*, 7(1), 95. doi:10.1186/1471-213X-7-95
- Les races bovines. (s. d.). *La-Viande.fr*. Repéré à https://www.la-viande.fr/animal-elevage/boeuf/racesbovines
- Li, X., Yang, L., Zhang, X. et Jiao, X. (2019). Prediction of Protein-Protein Interactions Based on Domain. Computational and Mathematical Methods in Medicine, 2019(2), 1-7. doi:10.1155/2019/5238406
- Lim, D., Kim, N.-K., Lee, S., Park, H., Cho, Y., Chai, H. et Kim, H. (2014). Characterization of Genes for Beef Marbling Based on Applying Gene Coexpression Network. *International Journal of Genomics*, 2014, 1-10. doi:10.1155/2014/708562
- Liptak, T. (1958). On the combination of independent tests. Magyar Tudomanyos Akademia Matematikai Kutato Intezenek Kozlomenyei.
- Loughin, T. M. (2004). A systematic comparison of methods for combining p-values from independent tests. *Computational Statistics and Data Analysis*, 47(3), 467-485. doi:10.1016/j.csda.2003.11.020
- Louveau, I., Perruchot, M. H., Bonnet, M. et Gondret, F. (2016). Invited review: Pre- and postnatal adipose tissue development in farm animals: From stem cells to adipocyte physiology. *Animal*, 10(11), 1839-1847. doi:10.1017/S1751731116000872
- Macaskill, P., Walter, S. D. et Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, 20(4), 641-654. doi:10.1002/sim.698
- Maison, P. (2010). La méta-analyse sur données résumées. *Recherche en soins infirmiers*, *101*, 18-24. Repéré à https://www.cairn.info/revue-recherche-en-soins-infirmiers-2010-2-page-18.htm#
- Mao, Y., Hopkins, D. L., Zhang, Y., Li, P., Zhu, L., Dong, P., ... Luo, X. (2016). Beef quality with different intramuscular fat content and proteomic analysis using isobaric tag for relative and absolute quantitation of differentially expressed proteins. *Meat Science*, *118*, 96-102. doi:10.1016/j.meatsci.2016.03.028
- Marot, G. (2015). Package ' metaMA '. Repéré à https://cran.rproject.org/web/packages/metaMA/metaMA.pdf
- Marot, G., Foulley, J.-L., Mayer, C.-D. et Jaffrezic, F. (2009). Moderated effect size and P-value combinations for microarray meta-analyses. *Bioinformatics*, 25(20), 2692-2699. doi:10.1093/bioinformatics/btp444
- Marot, G., Jaffrezic, F. et Rau, A. (2013). metaRNASeq : Differential meta-analysis of RNA-seq data. *R*, 1-10.
- Meunier, B., Bonnet, M., Picard, B., Ortigues, I., Agabriel, J., De La Torre Capitan, A. Sepchat, B. et Micol, D. (2014). *L'imagerie appliquée a la mesure de composition en tissus de la 6ème côte du bovin*. Clermont-Ferrand : 15èmes Journées Sciences du Muscle et Technologies des Viandes.
- Miller, M. F., Cross, H. R., Lunt, D. K. et Smith, S. B. (1991). Lipogenesis in acute and 48-hour cultures of bovine intramuscular and subcutaneous adipose tissue explants. *Journal of animal science*, 69(1), 162-170. doi:10.2527/1991.691162x
- Moisá, S. J., Shike, D. W., Graugnard, D. E., Rodriguez-Zas, S. L., Everts, R. E., Lewin, H. A., ... Loor, J. J. (2013). Bioinformatics analysis of transcriptome dynamics during growth in Angus cattle longissimus muscle. *Bioinformatics and Biology Insights*, 7, 253-270. doi:10.4137/BBI.S12328
- Monteils, V., Sibra, C., Ellies-Oury, M. P., Botreau, R., De la Torre, A. et Laurent, C. (2017). A set of indicators to better characterize beef carcasses at the slaughterhouse level in addition to the EUROP system. *Livestock Science*, 202(May), 44-51. doi:10.1016/j.livsci.2017.05.017
- Moreews, F., Rauffet, G., Dehais, P. et Klopp, C. (2011). SigReannot-mart: A query environment for expression microarray probe re-annotations. *Database*, 2011(June), 2-7. doi:10.1093/database/bar025
- Morgan, J. B., Wheeler, T. L., Koohmaraie, M., Crouse, J. D. et Savell, J. W. (1993). Effect of castration on myofibrillar protein turnover, endogenous proteinase activities, and muscle growth in bovine skeletal muscle. *Journal of animal science*, *71*(2), 408-414. doi:10.2527/1993.712408x
- Mourot, B. P., Gruffat, D., Durand, D., Chesneau, G., Prache, S., Mairesse, G. et Andueza, D. (2014). New approach to improve the calibration of main fatty acids by near-infrared reflectance spectroscopy in ruminant meat, (2011).
- Mudholkar, G. et George, E. (1979). The logit method for combining probabilities. Symposium on Optimizing Methods in Statistics, (Academic P; édité par J. Rustagi (Ed.)). New York.
- Naganathan, G. K., Grimes, L. M., Subbiah, J., Calkins, C. R., Samal, A. et Meyer, G. E. (2008). Visible/near-infrared hyperspectral imaging for beef tenderness prediction. *Computers and Electronics in Agriculture*, 64(2), 225-233. doi:10.1016/j.compag.2008.05.020
- Nie, L., Wu, G., Culley, D. E., Scholten, J. C. M. et Zhang, W. (2007). Integrative analysis of transcriptomic and proteomic data: Challenges, solutions and applications. *Critical Reviews in Biotechnology*, 27(2), 63-75. doi:10.1080/07388550701334212
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., ... Hermjakob, H. (2014). The MIntAct project IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*, *42*(D1), 358-363. doi:10.1093/nar/gkt1115
- Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bridge, A., ... Raghunath, A. (2012). Exchange Consortium (IMEx). *Nature Methods*, 9(4), 345-350. doi:10.1038/nmeth.1931.Protein
- Pallafacchina, G., Blaauw, B. et Schiaffino, S. (2013). Role of satellite cells in muscle growth and maintenance of muscle mass. *Nutrition, Metabolism and Cardiovascular Diseases*, 23(SUPPL1), 1-7. doi:10.1016/j.numecd.2012.02.002
- Park, S. J., Beak, S. H., Jung, D. J. S., Kim, S. Y., Jeong, I. H., Piao, M. Y., ... Baik, M. (2018). Genetic, management, and nutritional factors affecting intramuscular fat deposition in beef cattle - A review. Asian-Australasian Journal of Animal Sciences. doi:10.5713/ajas.18.0310
- Park, S. J., Kang, H. J., Na, S., Lee, S. H. et Baik, M. (2018). Differential expression of extracellular matrix and integrin genes in the longissimus thoracis between bulls and steers and their association with intramuscular fat contents. *Meat Science*, 136(October 2017), 35-43. doi:10.1016/j.meatsci.2017.10.008
- Pearson, K. (1904). Report on Certain Enteric Fever Inoculation Statistics. *British medical journal*, 2(2288), 1243-6. Repéré à http://www.ncbi.nlm.nih.gov/pubmed/20761760
- Pearson, K. (1933). On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn from a Parent Population Having a Known Probability Integral has Probably Been Drawn at Random. *Biometrika*, 25(3/4), 379. doi:10.2307/2332290
- Peng, Y. et Dhakal, S. (2015). Optical methods and techniques for meat quality inspection. *Transactions* of the ASABE, 58(5), 1371-1386. doi:10.13031/trans.58.11004
- Perez-Iratxeta, C. et Andrade, M. A. (2005). Inconsistencies over time in 5% of NetAffx probe-to-gene annotations. *BMC Bioinformatics*, 6, 2004-2006. doi:10.1186/1471-2105-6-183
- Perry, D., Yeates, A. et McKiernan, W. (1993). Meat yield and subjective muscle scores in medium

weight steers. Australian Journal of Experimental Agriculture, 33(7), 825. doi:10.1071/EA9930825

- Petersen, T. N., Brunak, S., Von Heijne, G. et Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785-786. doi:10.1038/nmeth.1701
- Pethick, D., Barendse, W., Hocquette, J., Thompson, J. et Wang, Y. (2007). Regulation of marbling and body composition growth and development, gene markers and nutritional biochemistry. Dans I. Ortigues-Marty (dir.), *Energy and Protein Metabolism and Nutrition* (p. 75–88). Wageningen, The Netherlands. : Wageningen Academic Publishers.
- Picard, B., Gagaoua, M., Al Jammas, M. et Bonnet, M. (2019). Beef tenderness and intramuscular fat proteomic biomarkers: Effect of gender and rearing practices. *Journal of Proteomics*, 200(February), 1-10. doi:10.1016/j.jprot.2019.03.010
- Picard, B., Lefaucheur, L., Berri, C. et Duclos, M. J. (2002). Muscle fibre ontogenesis in farm animal species. *Reproduction Nutrition Development*, 42(5), 415-431. doi:10.1051/rnd:2002035
- Pisani, D. F., Djedaini, M., Beranger, G. E., Elabd, C., Scheideler, M., Ailhaud, G. et Amri, E. Z. (2011). Differentiation of human adipose-derived stem cells into « brite » (brown-in-white) adipocytes. *Frontiers in Endocrinology*, 2(NOV), 1-9. doi:10.3389/fendo.2011.00087
- Polanin, J. R., Hennessy, E. A. et Tanner-Smith, E. E. (2017). A Review of Meta-Analysis Packages in
 R. Journal of Educational and Behavioral Statistics, 42(2), 206-242.
 doi:10.3102/1076998616674315
- Poleti, M. D., Regitano, L. C. A., Souza, G. H. M. F., Cesar, A. S. M., Simas, R. C., Silva-Vignato, B., ... Coutinho, L. L. (2018). Longissimus dorsi muscle label-free quantitative proteomic reveals biological mechanisms associated with intramuscular fat deposition. *Journal of Proteomics*, 179(November 2017), 30-41. doi:10.1016/j.jprot.2018.02.028
- Polkinghorne, R. J. et Thompson, J. M. (2010). Meat standards and grading. A world view. *Meat Science*, 86(1), 227-235. doi:10.1016/j.meatsci.2010.05.010
- Prieto, N., Ross, D. W., Navajas, E. A., Richardson, R. I., Hyslop, J. J., Simm, G. et Roehe, R. (2011). Online prediction of fatty acid profiles in crossbred Limousin and Aberdeen Angus beef cattle using near infrared reflectance spectroscopy. *Animal*, 5(1), 155-165. doi:10.1017/S1751731110001618
- Prieto, Nuria, Pawluczyk, O., Dugan, M. E. R. et Aalhus, J. L. (2017). A Review of the Principles and Applications of Near-Infrared Spectroscopy to Characterize Meat, Fat, and Meat Products. *Applied Spectroscopy*, 71(7), 1403-1426. doi:10.1177/0003702817709299
- Prior, R. L., Smith, S. B., Schanbacher, B. D. et Mersmann, H. J. (1983). Lipid metabolism in finishing bulls and steers implanted with oestradiol-17 β-dipropionate. *Animal Production*, 37(1), 81-88. doi:10.1017/S0003356100001562
- Qiao, X., Zhang, H. H., Liu, Y., Todd, M. J. et Marron, J. S. (2010). Weighted distance weighted discrimination and its asymptotic properties. *Journal of the American Statistical Association*, 105(489), 401-414. doi:10.1198/jasa.2010.tm08487
- Qin, L., Zhang, G., Cao, Y., Zhang, J., Zhao, Y. et Zhao, Z. (2011). Microarray Analysis on the Differences of Gene Expression in Longissimus Dorsi Muscle Tissue Between 1 and 24 Months Chinese Red Steppes. *Animal and Veterinary Advances*, 10(4), 428-436.
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria : R Foundation for Statistical Computing. Repéré à https://www.r-project.org/
- Rajesh, R. V., Heo, G. N., Park, M. R., Nam, J. S., Kim, N. K., Yoon, D., ... Lee, H. J. (2010). Proteomic analysis of bovine omental, subcutaneous and intramuscular preadipocytes during in vitro

adipogenic differentiation. Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics, 5(3), 234-244. doi:10.1016/j.cbd.2010.06.004

- Ramasamy, A., Mondry, A., Holmes, C. C. et Altman, D. G. (2008). Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. *PLoS Medicine*, 5(9), e184. doi:10.1371/journal.pmed.0050184
- Rice, P., Longden, I. et Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *The European Molecular Biology Open Software Suite*, *16*(6), 2-3.
- Rifai, N., Gillette, M. A. et Carr, S. A. (2006). Protein biomarker discovery and validation : the long and uncertain path to clinical utility, 24(8), 971-983. doi:10.1038/nbt1235
- Riley, R. D., Simmonds, M. C. et Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology*, 60(5), 431.e1-431.e12. doi:10.1016/j.jclinepi.2006.09.009
- Robelin, J. (1981). Cellularity of bovine adipose tissues: Developmental changes from 15 to 65 percent mature weight. *Journal of Lipid Research*, 22(3), 452-457.
- Robelin, J. (1986). Composition corporelle des bovins : Evolution au cours du développement et différence entre races. Université de Clermont-Ferrand II.
- Robelin, J. et Casteilla, L. (1990). Différenciation, croissance et développement du tissu adipeux. *INRA Productions Animales*, *1*(1), 1-3.
- Robelin, J., Geay, Y., Jailler, R. et Cuylle, G. (1975). ESTIMATION DE LA COMPOSITION DES CARCASSES DE JEUNES BOVINS A PARTIR DE LA COMPOSITION D'UN MORCEAU MONOCOSTAL PRÉLEVÉ AU NIVEAU DE LA 11e CÔTE.
- Roche, F. M., Hokamp, K., Acab, M., Babiuk, L. A., Hancock, R. E. W. et Brinkman, F. S. L. (2004). ProbeLynx: A tool for updating the association of microarray probes to genes. *Nucleic Acids Research*, 32(WEB SERVER ISS.), 471-474. doi:10.1093/nar/gkh452
- Roehe, R., Ross, D., Duthie, C.-A., Lambe, N., Anderson, C., Broadbent, C., ... Glasbey, C. (2013). Research Towards an Integrated Measurement of Meat Eating Quality (IMEQ). Scottish Government and Quality Meat Scotland (QMS), 2013(February 2014). Repéré à https://www2.gov.scot/Topics/farmingrural/Agriculture/Livestock/Meat/Beef/classification/IME QReport
- Romao, J. M., Jin, W., He, M., Mcallister, T. et Guan, L. L. (2013). Elucidation of Molecular Mechanisms of Physiological Variations between Bovine Subcutaneous and Visceral Fat Depots under Different Nutritional Regimes, 8(12), 1-14. doi:10.1371/journal.pone.0083211
- Rung, J. et Brazma, A. (2013). Reuse of public genome-wide gene expression data. *Nature Reviews Genetics*, 14(2), 1-11. doi:10.1038/nrg3394
- Russell, W. M. S. et Burch, R. L. (s. d.). La règle des 3 R : réduire, raffiner, remplacer. Repéré à https://www.inserm.fr/professionnels-recherche/recherche-pre-clinique/experimentation-animale/regle-3-r-reduire-raffiner-remplacer
- Ryu, S. Y. et Wendt, G. A. (2019). MetaMSD: meta analysis for mass spectrometry data. *PeerJ*, 7, e6699. doi:10.7717/peerj.6699
- Sadkowski, T., Ciecierska, A., Majewska, A., Oprz, J., Dasiewicz, K., Ollik, M., ... Motyl, T. (2014). Transcriptional background of beef marbling — Novel genes implicated in intramuscular fat deposition, 97(3), 32-41. doi:10.1016/j.meatsci.2013.12.017
- Schoonmaker, J. P., Cecava, M. J., Fluharty, F. L., Zerby, H. N. et Loerch, S. C. (2004). Effect of source and amount of energy and rate of growth in the growing phase on performance and carcass characteristics of early- and normal-weaned steers. *Journal of Animal Science*, 82(1), 273-282.

- Seale, P., Bjork, B., Yang, W., Kajimura, S., Chin, S., Kuang, S., ... Spiegelman, B. M. (2008). PRDM16 controls a brown fat/skeletal muscle switch. *Nature*, 454(7207), 961-967. doi:10.1038/nature07182
- Shabalin, A. A., Tjelmeland, H., Fan, C., Perou, C. M. et Nobel, A. B. (2008). Merging two geneexpression studies via cross-platform normalization. *Bioinformatics*, 24(9), 1154-1160. doi:10.1093/bioinformatics/btn083
- Shannon, P. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498-2504. doi:10.1101/gr.1239303
- Shen, Y. N., Kim, S. H., Yoon, D. H., Lee, H. G., Kang, H. S. et Seo, K. S. (2012). Proteome analysis of bovine Longissimus dorsi muscle associated with the marbling score. *Asian-Australasian Journal of Animal Sciences*, 25(8), 1083-1088. doi:10.5713/ajas.2012.12127
- Shingu, H., Hodate, K., Kushibiki, S., Ueda, Y., Watanabe, A., Shinoda, M. et Matsumoto, M. (2001). Profiles of growth hormone and insulin secretion, and glucose response to insulin in growing Japanese Black heifers (beef type): Comparison with Holstein heifers (dairy type). Comparative Biochemistry and Physiology - C Toxicology and Pharmacology, 130(2), 259-270. doi:10.1016/S1532-0456(01)00249-6
- Siezen, R. J. et Van Hijum, S. A. F. T. (2010). Genome (re-)annotation and open-source annotation pipelines. *Microbial Biotechnology*, *3*(4), 362-369. doi:10.1111/j.1751-7915.2010.00191.x
- Skalnikova, H., Motlik, J., Gadher, S. J. et Kovarova, H. (2011). Mapping of the secretome of primary isolates of mammalian cells, stem cells and derived cell lines. *Proteomics*, 11(4), 691-708. doi:10.1002/pmic.201000402
- Smith, S. B., Blackmon, T. L., Sawyer, J. E., Miller, R. K., Baber, J. R., Morrill, J. C., ... Wickersham, T. A. (2018). Glucose and acetate metabolism in bovine intramuscular and subcutaneous adipose tissues from steers infused with glucose, propionate, or acetate. *Journal of Animal Science*, 96(3), 921-929. doi:10.1093/jas/sky017
- Smith, S. B. et Crouse, J. D. (1984). Relative contributions of acetate, lactate and glucose to lipogenesis in bovine intramuscular and subcutaneous adipose tissue. *Journal of Nutrition*, *114*(4), 792-800. doi:10.1093/jn/114.4.792
- Smith, T.-C., Marcucci, M., Nolan, S. J., Iorio, A., Sudell, M., Riley, R., ... Williamson, P. R. (2016). Individual participant datameta-analyses compared with meta-analyses based on aggregate data (Review). *Cochrane Database of Systematic Reviews*, (9), MR000007. doi:10.1002/14651858.MR000007.pub4
- Spalding, K. L., Arner, E., Westermark, P. O., Bernard, S., Buchholz, B. A., Bergmann, O., ... Arner, P. (2008). Dynamics of fat cell turnover in humans. *Nature*, 453(7196), 783-787. doi:10.1038/nature06902
- Steen, R. W. J. et Kilpatrick, D. J. (1995). Effects of plane of nutrition and slaughter weight on the carcass composition of serially slaughtered bulls, steers and heifers of three breed crosses. *Livestock Production Science*, 43(3), 205-213. doi:10.1016/0301-6226(95)00046-N
- Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams, R. M., J. (1949). *The American soldier: Adjustment during army life (Studies in social psychology in World War II)*. Oxford, England: Princeton Univ. Press.
- Sullivan, G. M. et Feinn, R. (2012). Using Effect Size—or Why the P Value Is Not Enough . *Journal* of Graduate Medical Education, 4(3), 279-282. doi:10.4300/jgme-d-12-00156.1
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., ... Von Mering, C. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, 45(D1), D362-D368. doi:10.1093/nar/gkw937

- Taga, H., Chilliard, Y., Meunier, B., Chambon, C., Picard, B., Zingaretti, M. C., ... Bonnet, M. (2012). Cellular and molecular large-scale features of fetal adipose tissue: Is bovine perirenal adipose tissue Brown? *Journal of Cellular Physiology*, 227(4), 1688-1700. doi:10.1002/jcp.22893
- Tait, R. G. (2016). Ultrasound Use for Body Composition and Carcass Quality Assessment in Cattle and Lambs. Veterinary Clinics of North America - Food Animal Practice, 32(1), 207-218. doi:10.1016/j.cvfa.2015.09.007
- TERRE-NET. (s. d.). Marchés de la viande. *web-agri*. Repéré à http://www.web-agri.fr/gros-bovins-boucherie/334
- Thornton, K. J., Chapalamadugu, K. C., Eldredge, E. M. et Murdoch, G. K. (2017). Analysis of Longissimus thoracis Protein Expression Associated with Variation in Carcass Quality Grade and Marbling of Beef Cattle Raised in the Pacific Northwestern United States. Journal of Agricultural and Food Chemistry, 65(7), 1434-1442. doi:10.1021/acs.jafc.6b02795
- TIPPETT, L. H. C. (1931). The Methods of Statistics. London: Williams & Norgate Ltd.
- Tournayre, J., Cassar-Malek, I., Reichstadt, M., Picard, B., Kaspric, N. et Bonnet, M. (2015). Fat & MuscleDB: A database to understand tissue growth processes contributing to body or muscle composition. *Communication JOBIM*.
- Tseng, G. C., Ghosh, D. et Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Research*, 40(9), 3785-3799. doi:10.1093/nar/gkr1265
- Tseng, Y. H., Kokkotou, E., Schulz, T. J., Huang, T. L., Winnay, J. N., Taniguchi, C. M., ... Kahn, C. R. (2008). New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature*, 454(7207), 1000-1004. doi:10.1038/nature07221
- van den Berg, B. H. J., McCarthy, F. M., Lamont, S. J. et Burgess, S. C. (2010). Re-annotation is an essential step in systems biology modeling of functional genomics data. *PLoS ONE*, 5(5). doi:10.1371/journal.pone.0010642
- van den Berg, B. H., Konieczka, J. H., McCarthy, F. M. et Burgess, S. C. (2009). ArrayIDer: automated structural re-annotation pipeline for DNA microarrays. *BMC bioinformatics*, *10*, 30. doi:10.1186/1471-2105-10-30
- Vaudel, M., Verheggen, K., Csordas, A., Raeder, H., Berven, F. S., Martens, L., ... Barsnes, H. (2016). Exploring the potential of public proteomics data. *PROTEOMICS*, 16(2), 214-225. doi:10.1002/pmic.201500295
- Vernon, M. W., Beard, J. S., Graves, K. et Wilson, E. A. (1986). Classification of endometriotic implants by morphologic appearance and capacity to synthesize prostaglandin F**Supported in part by Biomedical Research Support grant RR 05374 from the Biomedical Research Support Branch, Division of Research Facilities and. *Fertility and Sterility*, 46(5), 801-806. doi:10.1016/s0015-0282(16)49814-6

Viechtbauer, W. (2019). metafor: Meta-Analysis Package for R, Version 2.1-0, (1), 271.

- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. et Snel, B. (2003). STRING: A database of predicted functional associations between proteins. *Nucleic Acids Research*, *31*(1), 258-261. doi:10.1093/nar/gkg034
- Walsh, C., Hu, P., Batt, J. et Santos, C. (2015). Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays*, 4(3), 389-406. doi:10.3390/microarrays4030389
- Wang, H. B., Zan, L. S. et Zhang, Y. Y. (2014). Profiling of the yak skeletal muscle tissue gene expression and comparison with the domestic cattle by genome array. *Animal*, 8(1), 28-35.

doi:10.1017/S1751731113001948

- Wang, Y. H., Bower, N. I., Reverter, A., Tan, S. H., De Jager, N., Wang, R., ... Lehnert, S. A. (2009). Gene expression patterns during intramuscular fat development in cattle. *Journal of Animal Science*, 87(1), 119-130. doi:10.2527/jas.2008-1082
- Williams, R. T. (2012). Using Robust Standard Errors to Combine Multiple Regression Estimates with Meta-Analysis (LOYOLA UNIVERSITY CHICAGO USING). Repéré à http://ecommons.luc.edu/cgi/viewcontent.cgi?article=1404&context=luc_diss
- Yancey, J. W. S., Apple, J. K., Meullenet, J. F. et Sawyer, J. T. (2010). Consumer responses for tenderness and overall impression can be predicted by visible and near-infrared spectroscopy, Meullenet-Owens razor shear, and Warner-Bratzler shear force. *Meat Science*, 85(3), 487-492. doi:10.1016/j.meatsci.2010.02.020
- Yates, B. Y. F. et Cochran, W. G. (2006). The Analysis of Groups, 162-187. doi:10.1002/0471746096.ch6
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., ... Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13), 1608-1615. doi:10.1093/bioinformatics/btq249
- Yuryev, A., Kotelnikova, E. et Daraselia, N. (2009). Ariadne's ChemEffect and Pathway Studio knowledge base. *Expert Opinion on Drug Discovery*, 4(12), 1307-1318. doi:10.1517/17460440903413488
- Zhang, B., Kirov, S. et Snoddy, J. (2005). WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research*, 33(SUPPL. 2), 741-748. doi:10.1093/nar/gki475
- Zhang, J., Zhong, C., Lin, H. X. et Wang, M. (2019). Identifying Protein Complexes from Dynamic Temporal Interval Protein-Protein Interaction Networks. *BioMed Research International*, 2019, 1-17. doi:10.1155/2019/3726721
- Zhang, Q., Lee, H. G., Han, J. A., Kim, E. B., Kang, S. K., Yin, J., ... Choi, Y. J. (2010). Differentially expressed proteins during fat accumulation in bovine skeletal muscle. *Meat Science*, *86*(3), 814-820. doi:10.1016/j.meatsci.2010.07.002
- Zhang, Y.-Y., Zan, L.-S., Wang, H.-B., Qing, L., Wu, K.-X., Quan, S.-A., ... Wang, C.-J. (2011). Differentially expressed genes in skeletal muscle tissues from castrated Qinchuan cattle males compared with those from intact males. *Livestock Science*, *135*(1), 76-83. doi:10.1016/j.livsci.2010.06.070
- Zhao, Y. M., Basu, U., Dodson, M. V., Basarb, J. A. et Guan, L. L. (2010). Proteome differences associated with fat accumulation in bovine subcutaneous adipose tissues. *Proteome Science*, *8*, 1-11. doi:10.1186/1477-5956-8-14
- Zintzaras, E. et Ioannidis, J. P. A. (2008). Meta-analysis for ranked discovery datasets: Theoretical framework and empirical demonstration for microarrays. *Computational Biology and Chemistry*, 32(1), 39-47. doi:10.1016/j.compbiolchem.2007.09.003
- Zollinger, A., Davison, A. C. et Goldstein, D. R. (2015). Meta-analysis of incomplete microarray studies. *Biostatistics*, 16(4), 686-700. doi:10.1093/biostatistics/kxv014