



HAL
open science

Conception d'un web service pour la fouille de données de génomique : application à la caractérisation de la myogenèse et de l'adipogenèse

Nicolas Kaspric

► **To cite this version:**

Nicolas Kaspric. Conception d'un web service pour la fouille de données de génomique : application à la caractérisation de la myogenèse et de l'adipogenèse. Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2016. Français. NNT : 2016CLF22674 . tel-01343388

HAL Id: tel-01343388

<https://theses.hal.science/tel-01343388>

Submitted on 8 Jul 2016

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 SCIENCES DE LA VIE, SANTE,
AGRONOMIE et ENVIRONNEMENT*

N° d'ordre : 684

Thèse présentée à l'Université Blaise Pascal pour l'obtention du grade de Docteur d'Université
(Spécialité : Bioinformatique)

**Conception d'un web service pour la fouille de
données de génomique : application à la
caractérisation de la myogenèse et de
l'adipogenèse**

Présenté par **Nicolas KASPRIC**

Thèse dirigée par **Brigitte PICARD** et **Muriel BONNET**

Soutenue publiquement le 24 Février 2016 devant le jury composé de :

Christine BRUN, Directeur de recherches, CNRS, Marseille (Rapporteur)

Frédérique LISACEK, Directeur de recherches, SIB, Genève (Rapporteur)

Gillian BUTLER-BROWNE, Directeur de recherches, INSERM, Paris (Examineur)

François ENAULT, Maître de Conférences, Université Blaise Pascal (Examineur)

Muriel BONNET, Chargée de recherches, INRA, Theix (Examineur)

Brigitte PICARD, Directeur de recherches, INRA, Theix (Examineur)

Carine CAPEL, Docteur en Biologie, Apis-Gene, Institut de l'élevage, Aubière (Membre invité)

Laboratoire d'accueil

Centre de Recherche de Clermont-Ferrand / Theix

Unité Mixte de Recherche sur les Herbivores

Equipe Amuvi

63122 Saint-Genès Champanelle



VetAgro Sup



Remerciements

Après trois années à l'INRA, plus particulièrement dans l'équipe Amuvi, de nombreuses personnes ont participé, de près ou de loin, à la réussite de ce projet de thèse. Merci à tous !

Tout d'abord, je tiens à remercier tous les membres du jury, Professeurs, Maître de Conférences, Directeurs et Chargés de recherches, d'avoir accepté de consacrer une partie de leur temps précieux à la lecture et à l'évaluation de ce travail de thèse.

Je remercie mes investisseurs la Région d'Auvergne et APIS-GENE qui ont soutenu financièrement cette thèse.

Je ne pense pas pouvoir suffisamment remercier Mme Brigitte PICARD et Mme Muriel BONNET qui ont assuré la co-direction de ma thèse et sans qui rien n'aurait pu aboutir.

Un grand merci à Brigitte PICARD de m'avoir accueilli au sein de son équipe Amuvi. Merci de m'avoir tant enseigné durant trois ans et de m'avoir expliqué les différents stades de la myogénèse. Merci pour tes qualités pédagogiques et humaines, et pour tous tes encouragements. Je suis heureux d'avoir pu te donner autant de lecture durant tous tes voyages.

Un grand merci à Muriel BONNET car je ne sais pas comment tu as pu supporter mes maladresses, oublis et mots poubelles durant un peu plus de trois ans. Je ne pourrai pas exprimer assez de gratitude pour tout le temps que tu as passé à m'apprendre à devenir un chercheur compétent, même s'il reste encore du travail. Je promets de te rembourser ces trois années lorsque j'aurai suffisamment appris de choses pour pouvoir t'en enseigner. Je sais que ce ne sera pas facile... Merci à toutes les deux d'avoir eu la patience de m'expliquer chaque infime partie du sujet. Merci de votre soutien et votre gentillesse tout au long de ces trois ans de travail.

Je souhaite exprimer ma gratitude à Mme Isabelle VEISSIER, Directrice de l'Unité Mixte de Recherche sur les Herbivores, pour m'avoir accueilli au sein de l'Unité. Qu'elle trouve ici le témoignage de ma reconnaissance.

Je souhaite remercier tout particulièrement M Axel POULET pour tous ces samedis passés à lire des BD, boire du thé et fumer des chichas. Je serai déjà hors course sans ces moments de détente et tes blagues. Je ne sais pas si je pourrai te remercier de tes encouragements et ton aide durant toutes ces années depuis le master. BAZINGA ! Merci à Mme Ludivine POULET qui a su me motiver avec mon surnom de « tête de mort » !

Je remercie M Matthieu REICHSTADT pour son aide inestimable en matière de programmation. Sans toi ProteINSIDE ne serait pas ce qu'il est. Merci pour ta bonne humeur constante, malgré toutes ces « Reichstadts » ! Tu restes un élément indispensable à la réussite de ce projet.

Merci à Jérémy TOURNAYRE, Jérémy HUANT et Gaëtan ROUMEAU, vous avez tenu bon face à mes chansons et blagues nulles ! Merci Jérémy T, car sans toi nous n'aurions pas de réseaux de PPI

présentables. Merci de m'avoir accompagné durant ces trois ans. Merci à Jérémy H pour les manip en labo et les mercredis JDR. Merci à Gaëtan pour toutes ses aides en programmation et sa bonne humeur. Merci pour les repas fort gustatifs au Toucan avec Matthieu, je pense qu'on a pris quelques kilos de tissu adipeux durant ces trois ans.

Je remercie enfin toute l'équipe Amuvi pour son accueil chaleureux, merci à : Isabelle CASSAR-MALEK, David CHADEYRON, Bruno MEUNIER, Vincent LARGEAU, Geneviève GENTES, Nicole DUNOYER, Anne-Sophie BAGE, Gonzalo CANTALAPIEDRA, Céline COSSOUL, Denis DURAND, Dominique GRUFFAT, Isabelle ORTHIGUES-MARTY, Agnès THOMAS, Jean VERNET, Arnaud DELAUAUD, Anne LISTRAT, Sylvie BARDOU-VALETTE, Jean-Luc MONTEL et Didier MICOL.

Merci à Christiane BARBOIRON pour toutes les heures passées à faire des gels 2D. Merci pour ta bonne humeur constante malgré les centaines de spots à récupérer, souvent avec la foi ! Merci Marie-Jo pour ta bonne humeur, ton écoute, ta gentillesse et ton aide précieuse chaque fois je t'ai sollicité.

Un grand merci à tous les stagiaires et thésards : Annabelle, Marion, Clothilde V, Julien, Benoît-Pierre, Maraw, Arnaud, Sarah B, Sara M, Mohammed, Julie, Refka, Hania, Amélie F, Amélie M, Kévin, Lahlou, Yunlong, Malek, Clothilde L et Valérie.

Merci à Christophe Chambon, de la plateforme « Exploration du Métabolisme : des gènes aux nutriments » du centre INRA de Theix, pour son expertise lors de l'identification des protéines.

Merci à toute l'équipe GAME qui m'a accompagné pour chaque dossier administratif et qui m'a permis d'effectuer cette thèse dans des conditions optimales. Merci à Nicolas BARRIERE, Maryse TEYSSEYDRE, Annie ROYET, Nathalie GOCEL, et Nicole CHONION.

Je souhaite enfin remercier toute ma famille qui m'a accompagné depuis le début de mes études. Sans vous je ne serais pas allé aussi loin, car beaucoup de professeurs n'auraient pas parié sur moi. Vous avez su me motiver et me donner plus que ce qu'il n'en faut pour réussir. Merci à mes parents, à ma sœur Natacha, à son mari Jonnathan et à leur fils Lucas (Waka), je vous aime tout simplement. Je ne sais pas comment remercier ma future femme Elodie qui m'accompagne depuis plus de trois ans et qui rend ma vie plus belle chaque jour. Nous la partagerons bientôt à trois. J'ai beaucoup de chance de vous avoir tous à mes côtés.

Curriculum vitae scientifique

Publications scientifiques

1 - Kaspric N, Picard B, Reichstadt M, Tournayre J, and Bonnet M. ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. May 2015. PLoS ONE. doi:10.1371/journal.pone.0128086

2 - Kaspric N, Reichstadt M, Picard B, Tournayre J, and Bonnet M. Protein function easily investigated by genomics data mining using the ProteINSIDE online tool. June 2015. Genomic and Computational Biology. doi: 10.18547/gcb.2015.vol1.iss1.e16

3 - Kaspric N, Picard B, Vonnahme K, Chambon C and Bonnet M. How mapping the muscle and adipose tissues proteomes could help in the understanding of the “adipose-muscular” cross-talk during fetal growth? *En préparation pour le journal Molecular and Cellular Proteomics*

Communications à des congrès nationaux et internationaux

1 - Bonnet M., Kaspric N., and Picard B. Quest for biomarkers of the lean-to-fat ratio by Proteomics in beef production. Proceeding of the 3rd meeting of working groups 1, 2, 3 of COST Action FA1002, Kosice, Slovaquie, April 25-26 2013. p43-44 (et communication orale).

2 - Kaspric N., Reichstadt M., Picard B., Tournayre J., and Bonnet M. ProteINSIDE: a web service to computerize an in-depth analysis of functions, sequences, secretions and interactions for proteins. RECOMB Comparative Genomics (Research in Computational Molecular Biology), Lyon, France, October 17-18-19 2013. p11 (et communication affichée).

3 - Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Protein function easily investigated by genomics data mining using the ProteINSIDE web service. Proceeding of the 2nd International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO), Granada, Spain, April 7-9 2014, volume 1, p890-902 (et communication orale).

4 - Tournayre J., Cassar-Malek I., Reichstadt M., Picard B., **Kaspric N.,** Bonnet M. Fat and Muscle DB: integrating ‘omics’ data from adipose tissue and muscle. Proceeding of the 4rd meeting of working groups 1, 2, 3 of COST Action FA1002, Milano, Italie, November 16-18 2014 (et communication orale).

5 - Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Proteomics data from ruminants easily investigated using ProteINSIDE. Proceeding of the 4rd meeting of working groups 1, 2, 3 of COST Action FA1002, Milano, Italie, November 16-18 2014 (et communication affichée).

6 - Kaspric N., Reichstadt M., Picard B., Tournayre J., and Bonnet M. Proteome data mining using ProteINSIDE online tool. Journées Ouvertes en Biologie, Informatique & Mathématiques (JOBIM), Clermont-Ferrand, France, July 6-9 2015 (et communication orale).

7 - Tournayre J., Cassar-Malek I., Reichstadt M., Picard B., **Kaspric N.** and Bonnet M. Fat&MuscleDB: Integrating "omics" data from adipose tissue and muscle. Journées Ouvertes en Biologie, Informatique & Mathématiques (JOBIM), Clermont-Ferrand, France, July 6-9 2015 (et communication orale).

8 - Bonnet M, **Kaspric N.**, Vonnahme K., Barboiron C., Viala D., Chadeyron D. and Picard B. Maternal nutrient restriction followed by realimentation from early to mid-gestation in beef cows impacts proteomes of fetal muscular and adipose tissues. 61st International Congress of Meat Science and Technology, Clermont-Ferrand, France, 23-28 August 2015 (et communication affichée).

9 - Picard B., Barboiron C., **Kaspric N.**, Vonnahme K. and Bonnet M. Impact on muscle fiber properties of maternal nutrient restriction followed by realimentation from early to mid-gestation in beef cows. 61st International Congress of Meat Science and Technology, Clermont-Ferrand, France, 23-28 August 2015 (et communication affichée).

Communications orales dans le cadre de journées scientifiques

1 – Journée des doctorants et post-doctorants de l'UMRH 2013 (05/02/2013)

Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Identification des interactions moléculaires entre le tissu adipeux et le muscle chez le bovin par une approche bioinformatique. Centre INRA de Theix.

2 – Journée des doctorants et post-doctorants de l'UMRH 2014 (25/02/2014)

Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Identification des interactions moléculaires entre le tissu adipeux et le muscle chez le bovin par une approche de bioinformatique. Centre INRA de Theix.

3 – Journée des rencontres AuDACES 2014 (17/06/2014)

Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. L'étude des protéomes de ruminants facilitée par l'utilisation du web service ProteINSIDE. Campus des Cézeaux, Journée AuDACES (Auvergne Développements et Calcul en Environnement Scientifique).

4 – Journée des doctorants et post-doctorants de l'UMRH 2015 (24/02/2015)

Kaspric N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Identification des interactions moléculaires entre le tissu adipeux et le muscle chez le bovin par une approche de bioinformatique. Centre INRA de Theix.

Activités complémentaires

2013 : Enseignement (18h) dans le cadre de modules informatiques en licence biologique (IUT Blaise Pascal)

2014 : Organisation et animation de la journée de doctorants et post-doctorants de l'UMRH 2014 (conférence sur une journée durant laquelle chaque doctorant de l'UMRH a la possibilité de présenter ses travaux)

2014 : Article pour la revue interne Herbimel INRA de Theix « Faire sa thèse à l'INRA » (co-écrit avec Julie Peyrat)

2014–2015 : Enseignement (60h) dans le cadre de modules informatiques en licence biologique (IUT Blaise Pascal)

2015 : encadrement d'un stagiaire bioinformaticien Antoine Perot (10 semaines) pour la mise en place d'une passerelle entre plusieurs outils bioinformatiques.

2015 : Article pour la revue régionale Auvergne Sciences (<http://www.auvergnesciences.com/> ; *en attente de parution*)

Sommaire

INTRODUCTION ET ETUDE BIBLIOGRAPHIQUE.....	1
INTRODUCTION.....	3
ETUDE BIBLIOGRAPHIQUE	5
<i>1 – Caractéristiques structurales et métaboliques des tissus adipeux et musculaires ..</i>	<i>5</i>
1.1 – Le tissu musculaire	5
Les différents types de fibres du muscle squelettique adulte	6
La classification des fibres.....	7
La plasticité des fibres musculaires.....	8
1.2 – Les tissus adipeux	8
Les différents tissus adipeux	8
Les différentes fonctions des adipocytes.....	9
<i>2 – L’ontogenèse du muscle et du TA</i>	<i>11</i>
2.1 – Ontogenèse du muscle.....	11
La myogenèse chez les mammifères.....	12
La myogenèse chez le bovin.....	13
De 60 à 180 jpc : la prolifération et le début de la différenciation.....	14
De 180 à 260 jpc : la différenciation et l’acquisition des fonctions biologiques	16
La régulation de la myogenèse	18
2.2 – Ontogenèse des tissus adipeux	19
L’adipogenèse chez les mammifères	19
L’adipogenèse chez le bovin	22
De 110 à 180 jpc : prolifération et différenciation précoce des progéniteurs adipeux par une régulation du cycle cellulaire et de l’apoptose	22
De 180 à 260 jpc : augmentation du volume des adipocytes par un contrôle de la différenciation et de l’hypertrophie des cellules adipeuses	24
<i>3 – Interactions entre les tissus musculaires et adipeux</i>	<i>25</i>
3.1 – Des interactions pour la régulation du nombre de cellules musculaires et adipeuses	26
3.2 – Des interactions pour la régulation du volume des cellules musculaires et adipeuses	28
<i>4 – Intégration de données publiques pour une meilleure compréhension de l’adipogenèse, de la myogenèse et des interactions entre le TA et le muscle</i>	<i>30</i>
<i>5 – La fouille de données de génomique par bioinformatique.....</i>	<i>33</i>
5.1 – Disponibilité des données de génomique	33

5.2 – Stratégie d’analyse pour une interprétation fonctionnelle de listes de gènes ou de protéines	34
6 – Revue des outils bioinformatique pour l’analyse de données de génomique	36
6.1 – Les bases de données référençant les protéines et les gènes	36
Les bases de données biologiques en génomique	37
Cas des ruminants	41
6.2 – Ontologie des gènes : l’annotation fonctionnelle des gènes et des protéines	42
Utilisation avancée de la Gene Ontology	44
Cas des ruminants	46
6.3 – La prédiction des protéines sécrétées.....	47
L’étude du sécrétome	47
Cas des ruminants	50
6.4 – Les interactions de protéines et voies métaboliques.....	50
Les interactions protéine-protéine	50
Les bases de données d’interactions de protéines	52
Réseaux d’interactions de protéines.....	54
Cas des ruminants	56
7 – Les outils d’analyse intégrative disponibles	56
8 – Conclusion et objectifs du travail de thèse	58
MATERIELS ET METHODES.....	61
1 – Données expérimentales de protéomique.....	63
1.1 – Animaux et échantillons	63
Cinétique d’ontogenèse des tissus.....	63
Description des protéomes musculaires et adipeux à 140 jours de vie fœtale.....	64
1.2 – Protéomique.....	65
Extraction et identification des protéines de l’ontogenèse des tissus	65
1.3 – Analyses bioinformatiques et clustérisations	67
Données de l’ontogenèse des tissus (première expérimentation).....	67
Données des protéomes fœtaux à 140 jpc (seconde expérimentation)	68
2 – Conception de l’outil bioinformatique ProteINSIDE	68
2.1 – Développement des modules d’analyse	69
Module « ID Mapping »	70
Module « GO »	70
Module « Secreted Proteins ».....	71

Module « Protein Interaction ».....	71
2.2 – Développement web de ProteINSIDE.....	72
RESULTATS.....	75
<i>1 – Développement d’un outil de fouille de données de génomique dédié aux ruminants : ProteINSIDE.....</i>	<i>77</i>
1.1 – Introduction.....	77
1.2 – La version bêta de ProteINSIDE.....	78
Article 1.....	80
Conclusion de l’étude.....	95
1.3 – ProteINSIDE : mises à jour pour renforcer la validité des résultats.....	95
Article 2.....	99
Conclusion de l’étude.....	115
1.4 - Conclusion générale à propos de l’outil ProteINSIDE.....	115
<i>2 – Mise en application de ProteINSIDE.....</i>	<i>118</i>
2.1 – Introduction.....	118
Article 3.....	119
2.2 – Conclusion.....	144
<i>3 – Caractérisation du protéome fœtal du muscle et du tissu adipeux à 140 jpc.....</i>	<i>146</i>
3.1 – Impact de l’alimentation maternelle sur le protéome des tissus adipeux et musculaires.....	146
Introduction.....	146
Communication 4.....	148
Conclusion.....	153
3.2 – Description des protéomes adipeux et musculaire à 140 jpc.....	153
Introduction.....	153
Article 5.....	154
Conclusion.....	192
DISCUSSION GENERALE, CONCLUSION ET PERSPECTIVES.....	195
<i>1 – L’analyse intégrative de données de génomique chez les ruminants grâce à ProteINSIDE.....</i>	<i>197</i>
1.1 – Utilisation et valorisation de ProteINSIDE.....	197
1.2 – Limites à l’utilisation de ProteINSIDE.....	198
1.3 – Perspectives d’évolution de ProteINSIDE.....	199
<i>2 – Les protéomes musculaire et adipeux au cours de la croissance fœtale bovine.....</i>	<i>200</i>

2.1 – Apports de connaissances par l’analyse du protéome fœtal.....	200
Les fonctions biologiques nouvellement identifiées dans les protéomes musculaires et adipeux.....	201
Les protéines sécrétées par les tissus adipeux et musculaires et leur implication dans le dialogue muscle-TA	206
2.2 – Les limites de notre étude.....	207
3 – <i>Conclusions et application du travail de thèse</i>	209
ANNEXES.....	213
REFERENCES.....	247

Liste des figures et tableaux

Figures

Figure 1 : Organisation structurale d'un muscle squelettique. Il s'agit d'une coupe du muscle selon les différentes parties qui le compose (de la fibre et ses filaments jusqu'au tissu musculaire). Dans ce sens, le muscle est composé de nombreux faisceaux de fibres entourés par l'épymisium. Des nerfs, des vaisseaux sanguins et du tissu adipeux sont présents entre les faisceaux de fibres. Ces faisceaux sont eux-mêmes composés de nombreuses fibres elles-mêmes composées de nombreuses myofibrilles, elles-mêmes composées de nombreux myofilaments.

Figure 2 : Organisation structurale d'une fibre musculaire en microscopie électronique (Tortora et Grabowski, 1994). On définit un sarcomère comme étant le segment entre deux disques Z voisins (de l'allemand « zwischen », signifiant « entre »). Dans les coupes longitudinales des muscles en microscopie électronique, les disques Z apparaissent comme une série de lignes foncées. À côté d'un disque Z, on retrouve la bande I (pour isotropique). La bande A (pour anisotropique) vient ensuite. On retrouve une partie plus pâle dans cette région appelée la zone H (de l'allemand « heller », pour « pâle »). Finalement incluse dans cette dernière, la mince strie M (de l'allemand « mittel », signifiant « centre ») se situe à l'intérieur. Les bandes A et I ont été nommées d'après leurs propriétés sous microscope polarisant. Les bandes A et I, ainsi que la ligne Z, sont visibles au microscope électronique.

Figure 3 : Plasticité des fibres musculaires. Les fibres pures expriment une seule isoforme de MyHC, les hybrides expriment plusieurs MyHC simultanément

Figure 4 : Différences structurales entre les 3 types d'adipocytes (Nedergaard et Cannon, 2013).

Figure 5 : Principales étapes de la myogenèse.

Figure 6 : Evolution des isoformes de chaînes lourdes de myosine (MyHC) au cours de la vie fœtale chez le bovin (Picard et al., 2003) suivie à l'aide d'anticorps spécifiques

Figure 7 : Evolution des activités enzymatiques de la LDH et de l'ICDH au cours de la vie fœtale dans différents types de muscles (Gagnière et al., 1999). La différenciation métabolique des muscles est précoce et a lieu au cours de la vie fœtale. Le métabolisme des fibres est déterminé par les activités enzymatiques des fibres, notamment de l'isocitrate déshydrogénase (ICDH ; métabolisme oxydatif) et de la lactate déshydrogénase (LDH ; métabolisme glycolytique). Ces deux enzymes représentent des voies métaboliques qui produisent de l'ATP, l'énergie nécessaire à la contraction des muscles. Le métabolisme oxydatif est celui des fibres de type I et IIA qui produisent plus d'ATP par molécule de glucose que les fibres rapides à métabolisme glycolytique (type IIX et IIB ; Tableau 1).

Figure 8 : Synthèse regroupant différentes protéines identifiées ayant un rôle connu ou hypothétique dans la migration cellulaire, la prolifération des myoblastes, la sortie du cycle cellulaire conduisant à la fusion des myoblastes et leur différenciation en myotubes ou à leur mort par apoptose (adapté de (Chaze et al., 2008)). Nous n'avons pas mentionné les protéines impliquées dans la différenciation métabolique et contractile des fibres musculaires. Les étapes de la myogenèse telles que la migration des précurseurs et leur prolifération se déroulent en même temps. Il en est de même pour les étapes de fusion et de différenciation des myotubes des différentes générations.

Figure 9 : Principales étapes de l'adipogenèse.

Figure 10 : Nombre total d'adipocytes périrénaux, volume de ces adipocytes (pL) et poids du TA périrénal (g) à 110 (pour le poids seulement), 180, 210 et 260 jpc pour les fœtus des races Charolaise et Blonde d'Aquitaine (adapté de (Taga et al., 2012)). Les résultats sont des moyennes et des écarts types, A, B, A x B : effets significatifs de l'âge et de la race ($P < 0.001$) et des interactions entre l'âge et la race (à $P < 0.05$).

Figure 11 : Protéines potentiellement impliquées dans la prolifération des progéniteurs adipeux (adapté de la thèse d'Hager Taga, (Taga, 2011)). Les flèches vertes représentent les activations et les flèches rouges les inhibitions du cycle cellulaire. Les protéines en marron ont été identifiées par Taga et al. (Taga et al., 2012) et ont un rôle potentiel dans la prolifération des précurseurs adipocytaires par le contrôle de la régulation du cycle cellulaire, de l'apoptose ou du retardement de la différenciation induit par PPAR γ .

Figure 12 : Cellules à l'origine des lignages des précurseurs musculaires et adipeux suivant les activateurs de leur détermination cellulaire (Bonnet et al., 2015).

Figure 13: Les deux lignées de progéniteurs Myf5, dans chaque dépôt de TA, varient en fonction de la localisation anatomique (Sanchez-Gurmaches et Guertin, 2014). On y observe les TA blancs (WAT) et brun (BAT) de plusieurs origines, comme : le TA inguinal (aine ; ingWAT), rétropéritonéale (rénal ; rWAT), sous-scapulaire (sous l'épaule ; sBAT), interscapulaire (entre les épaules ; iBAT), périgonadal (entourant l'utérus et les ovaires chez les femmes, l'épididyme et les testicules chez les mâles ; pgWAT), mésentère (surface des intestin grêle ; imWAT) et sous-cutanée (asWAT).

Figure 14 : Stratégie générale d'analyse de listes de protéines ou gènes issus d'étude de protéomique (adapté de (Laukens et al., 2015)).

Figure 15 : (A) L'outil de visualisation KEGG Atlas est une interface graphique avancée pour explorer les cartographies de KEGG avec la capacité d'interagir avec les voies métaboliques pour en obtenir plus d'information. (B) L'outil de visualisation « Pathway Browser » de Reactome permet d'analyser des données et de cartographier les voies métaboliques ou d'enrichir les résultats avec d'autres voies disponibles dans la base.

Figure 16 : Arbre représentatif des différents niveaux de relations amenant jusqu'au terme GO « regulation of adipose tissue development ». Les couleurs de flèche correspondent aux différentes relations possibles entre les termes GO (<http://www.ebi.ac.uk/QuickGO/>).

Figure 17 : Arbre représentatif des différents niveaux de preuves des annotations par termes GO (<http://www.geneontology.org/>).

Figure 18 : Fiche descriptive du terme GO « glycolytic process » (glycolyse) provenant de la base de données AmiGO. La partie A synthétise les informations liées au terme GO (code, dénomination, définition, etc...). La partie B représente toutes les entités annotées par le terme GO (protéines, gènes, « Evidence code » de l'annotation, espèce, base de données source de l'information...).

Figure 19 : Résultats graphiques obtenus en utilisant l'outil agriGO. (A) Le graphique en barres représente les termes GO les plus généraux retrouvés dans l'analyse en fonction du pourcentage de gènes annotés (les barres bleues donnent accès aux gènes annotés et les barres vertes représentent les annotations du terme GO selon l'espèce d'étude). (B) agriGO représente aussi les résultats sous la forme d'un arbre de parenté, la couleur des termes représente les degrés d'enrichissement des termes (rouge pour « très enrichis » à blanc pour « non enrichis » ; les carrés donnent aussi accès aux gènes annotés).

Figure 20 : Représentation non exhaustive des méthodes de détection des interactions moléculaires répertoriées sur le site OLS (Ontology Lookup Service (Cote et al., 2010)).

Figure 21 : Liens entre les bases de données d'interactions moléculaires (modifié à partir de (Klingstrom et Plewczynski, 2011)) et leurs appartenances à des méta-bases de données d'interactions.

Figure 22 : Diagramme représentant le fonctionnement du workflow de l'outil Expander (Ulitsky et al., 2010). Les rectangles rouges indiquent les fichiers d'entrée, les rectangles bleus indiquent les différents modules d'analyse et les noms des algorithmes sont écrits en orange.

Figure 23 : Chronologie de la croissance des tissus musculaire et adipeux au cours de la vie fœtale chez le bovin. Les cellules musculaires prolifèrent jusqu'à fixation du nombre total de fibres musculaires (NTF), se différencient et acquièrent des fonctions biologiques. Les cellules adipeuses prolifèrent et se différencient continuellement, bien que la prolifération ait lieu essentiellement en période périnatale (pour revue (Bonnet et al., 2010; Picard et al., 2002)).

Figure 24 : Protocole nutritionnel de la seconde expérimentation (Gonzalez et al., 2013).

Figure 25 : Schéma explicatif d'une électrophorèse bidimensionnelle. Dans un premier temps (1) les protéines migrent dans un gradient de pH et sont séparées en fonction de leur point isoélectrique (pH auquel la charge de la protéine est nulle) puis (2) elles sont séparées en fonction de leur poids moléculaire dans un gel SDS-PAGE. Après coloration au bleu de coomassie colloïdal G250 (3), les gels sont scannés et les spots protéiques sont analysés par analyse d'image.

Figure 26 : Différentes étapes du protocole d'identification des protéines suivi au laboratoire Amuvi. Après la séparation des protéines par gel 2D, celui-ci est scanné pour être analysé informatiquement afin de révéler les spots protéiques communs entre plusieurs gels. Les spots d'intérêt sont extraits et les protéines sont digérées par action de la trypsine. Les échantillons sont ensuite analysés par spectrométrie de masse et les résultats sont comparés à une base de données pour identifier les protéines.

Figure 27 : Analyse des données de protéomique de la myogenèse du muscle *Semitendinosus* de fœtus bovins de race charolaise (Chaze et al., 2008). (A) Analyse en composante principale illustrant la distinction des 5 stades étudiés (n=5 fœtus/stade) à partir de l'abondance des protéines révélées. (B) Analyse hiérarchique illustrant l'abondance de 120 protéines communes aux 5 stades étudiés et permettant la distinction de 5 profils d'évolution d'abondance de protéines différents (5 clusters). En vert les protéines peu abondantes, en rouge les protéines très abondantes.

Figure 28 : Organigramme des trois parties de ProteINSIDE : la base de données, l'interface web et le workflow. Ce dernier est divisé en quatre modules d'analyse.

Figure 29 : Les quatre modules initiaux de ProteINSIDE.

Figure 30 : Les méthodes de sélection des termes GO dans la version bêta et la version aboutie de ProteINSIDE sont comparées. (A) dans la version bêta, les termes GO reliés à l'état de sécrétion d'une protéine n'étaient pas forcément reliés à la sécrétion de protéines (termes encadrés en vert dans A). (B) les termes sont sélectionnés en partant du début de l'arborescence (encadrés en vert dans B), ainsi les termes qui n'étaient pas sélectionnés avant (terme en rouge dans B) sont pris en compte avec cette nouvelle méthode.

Figure 31 : Un clic sur une protéine donne des informations la concernant *via* la fiche UniProt et la prédiction de SignalP (si le module est utilisé). L'affichage d'un seul lien entre deux protéines améliore considérablement l'affichage.

Figure 32 : Nombre total des analyses effectuées sur ProteINSIDE au 16 décembre 2015. On peut y observer les différents pics d'activités du site selon les mois. Un diagramme montre la répartition des analyses effectuées pour toutes les espèces.

Figure 33 : Les fonctions biologiques les plus enrichies dans le protéome total et commun aux tissus adipeux et musculaires fœtaux.

Figure 34 : Les fonctions biologiques les plus enrichies dans le protéome spécifique au tissu musculaire fœtal.

Figure 35 : Les fonctions biologiques les plus enrichies dans le protéome spécifique du tissu adipeux fœtal.

Tableaux

Tableau 1 : Classement des différents types de fibres musculaires en fonction de leurs propriétés métaboliques et contractiles.

Tableau 2 : Différences structurales et fonctionnelles entre les 3 types d'adipocytes (adapté de (Park et al., 2014)).

Tableau 3 : Recensement des protéines et transcrits identifiés pour 107 GSE et 183 publications scientifiques en rapport avec l'adipogenèse ou la myogenèse (novembre 2015). Ce recensement a été réalisé grâce à l'outil Fat&MuscleDB (Tournayre et al., 2015) pour des expérimentations *in vitro* et *in vivo* chez le bovin, l'homme, le rat, la souris, le porc et le mouton.

Tableau 4 : Référencement et description générale des bases de données biologiques utilisées communément (nom de la base de données, utilisation gratuite, description, données répertoriées, qualité des mises à jour et dernière publication).

Tableau 5 : Table de contingence deux à deux utilisée pour calculer l'enrichissement d'un terme GO qui annote une liste Z de gènes par rapport à la liste X des gènes de l'espèce.

Tableau 6 : Les outils les plus communément utilisés pour la prédiction de protéines potentiellement sécrétées grâce à un peptide signal ou par d'autres voies de sécrétion (prédiction de la localisation cellulaire). Pour chaque outil, les auteurs ont calculé les taux de faux positifs liés à leur analyse qui sont reportés ici.

Tableau 7 : Listes des principales bases de données d'interactions moléculaires représentées dans la Figure 21. Sont rapportés les liens vers les sites internet des bases, le nombre d'interactions référencées en octobre 2015 et aussi les types d'interactions : PPI (protéine-protéine interaction), NPI (acide-nucléique-protéine interaction), SPI (petite molécule-protéine interaction), CURE (curation interne des interactions), IMEx et MIMIx (pour l'appartenance à un consortium de PPI), et PRED (prédiction d'interactions).

Tableau 8 : Résultats produits par la version bêta et la version aboutie de ProteINSIDE.

Liste des abréviations

AG : Acide Gras

CSE : Cellule Souche Embryonnaire

CSM : Cellule Souche Mésenchymateuse

EBI : European Bioinformatics Institute

GEO : Gene Expression Omnibus

GO : Gene Ontology

ID : Identifiant

jpc : jour post-conception

KEGG : Kyoto Encyclopedia of Genes and Genomes

LT : *Longissimus Thoracis*

MyHC : Myosin Heavy Chain (chaîne lourde de myosine)

NCBI : National Center for Biotechnology Information

NTF : Nombre Total de Fibres

PPI : Protein-Protein Interaction

ST : *Semitendinosus* (muscle)

TA : Tissu Adipeux

TG : TriGlycérides

INTRODUCTION ET ETUDE BIBLIOGRAPHIQUE

Introduction

Pour les espèces agricoles à viande, c'est la qualité musculaire de l'animal qui représente sa valeur bouchère. Les animaux sont classés en fonction de leur état d'engraissement et de la conformation de leur carcasse. Des chiffres désignent le niveau d'engraissement ou d'adiposité de l'animal (de 1 à 5, 1 indique une carcasse très maigre et 5 une carcasse très grasse). Le classement E.U.R.O.P classent par des lettres les carcasses selon leur conformation qui rend donc compte du développement musculaire de l'animal (E pour une conformation excellente et P pour une conformation médiocre). À l'abattoir, ce classement détermine le prix de vente de la carcasse et rend compte de la masse de muscle/viande qui sera extraite de la carcasse, mais également des qualités sensorielles potentielles de la viande. La valeur économique de la carcasse d'un animal résulte donc de l'équilibre entre la masse de tissus musculaire et adipeux qui la composent. Dans un contexte d'augmentation de la population mondiale consommatrice de viande, une stratégie pour augmenter la masse, ou le volume de viande produite, est d'augmenter l'efficacité productive des bovins en priorisant la croissance des muscles au détriment de celle des tissus adipeux (TA). À l'inverse, il peut y avoir une nécessité à augmenter le dépôt de lipides dans les TA musculaires des jeunes bovins de race à viande. En effet, l'abattage de bovins d'environ 15 à 19 mois pour produire une carcasse moyenne de 405 kg génère des viandes avec des teneurs en lipides souvent inférieures aux 3 à 4 % requis pour une bonne appréciation par les consommateurs. L'équilibre entre le développement du muscle et celui des TA est donc un paramètre de la biologie développementale à comprendre pour pouvoir influencer sur l'efficacité de production et/ou sur la qualité des viandes.

L'équilibre entre la masse des muscles relativement aux TA dépend du nombre et du volume des cellules musculaires relativement aux cellules adipeuses. Or, l'état actuel des connaissances indique que les mécanismes qui déterminent le nombre de cellules sont largement moins bien connus que ceux qui régulent leur volume. En outre, les connaissances actuelles suggèrent des interactions entre le TA et le muscle qui influeraient sur le nombre et/ou le volume selon les phases de croissance. Lors des phases précoces de la croissance, des interactions entre tissus orienteraient la détermination ou la différenciation de progéniteurs adipeux, myogéniques ou adipo-myogéniques, et influenceraient le nombre de cellules musculaires *versus* adipeuses. Tout au long de la croissance, des interactions entre tissus réguleraient le partage des nutriments entre les tissus et affecteraient le volume des cellules musculaires *versus* adipeuses. Or, chez les grands mammifères comme le bovin, les mécanismes qui influent sur le nombre de cellules ont lieu essentiellement pendant la vie fœtale. Aussi, le fœtus bovin est un bon modèle pour comprendre les mécanismes et les interactions entre tissus susceptibles de participer au déterminisme du nombre et du volume des cellules. Cependant, à ce jour, les données sur les caractéristiques cellulaires et moléculaires qui accompagnent la croissance des muscles et des TA chez le fœtus bovins restent limitées et il s'agit essentiellement de données de génomique. Or chez les ruminants la faible annotation et l'absence de ressources bioinformatiques dédiées sont des freins à l'analyse des données de génomiques. **Aussi, l'objectif de cette thèse est de concevoir un serveur web pour une fouille des données de génomique chez le bovin, afin d'améliorer les connaissances concernant les mécanismes moléculaires qui participent à la croissance (par hyperplasie et hypertrophie) des tissus musculaires et adipeux.**

Ce manuscrit est composé d'une étude bibliographique rapportant les mécanismes cellulaires et moléculaires qui accompagnent la croissance des tissus adipeux (adipogenèse) et musculaires (myogenèse) chez les mammifères, puis plus particulièrement chez les ruminants dont le bovin. Elle recense les données publiques qui pourraient aider à comprendre les mécanismes de la myogenèse, de l'adipogenèse et des interactions entre ces deux processus. L'utilisation de ce type de données suppose de pouvoir réaliser une fouille de données sur des milliers de gènes/protéines d'intérêt pour cette compréhension. C'est pourquoi une revue des outils bioinformatiques et bases de données biologiques disponibles pour la fouille de données de génomique est présentée. Il ne s'agit pas d'une revue exhaustive, mais d'une présentation des outils et des bases de données les plus couramment utilisés dans la littérature. Nous décrivons ensuite les dispositifs expérimentaux et les différentes analyses réalisés dans ce travail de thèse. Les résultats obtenus sont exposés en trois points : la production d'un service web pour la fouille de données de génomique chez le ruminant, l'évaluation des performances de cet outil comparativement à des outils existants dans d'autres espèces, et la production de connaissance sur des mécanismes qui accompagnent la croissance des tissus musculaires et adipeux. Enfin, une discussion et une conclusion générales résument les principaux acquis et les perspectives de ce travail.

Étude bibliographique

1 – Caractéristiques structurales et métaboliques des tissus adipeux et musculaires

Pour les bovins, la qualité de la carcasse et de la viande est le reflet des proportions en tissus musculaires et adipeux. Elles dépendent de la mise en place, mais aussi des caractéristiques structurales et fonctionnelles de ces tissus au cours des différentes étapes du développement.

1.1 – Le tissu musculaire

Le muscle squelettique est essentiellement composé de fibres musculaires (80 à 90% de la surface du muscle) entourées de tissu conjonctif (l'épymysium entoure le muscle, le pérmysium se situe autour des faisceaux de fibres et l'endomysium autour de chaque fibre ; [Figure 1](#)). Chez les mammifères, le muscle représente environ 55% du poids du corps adulte ([Bonnet et al., 2010](#)).

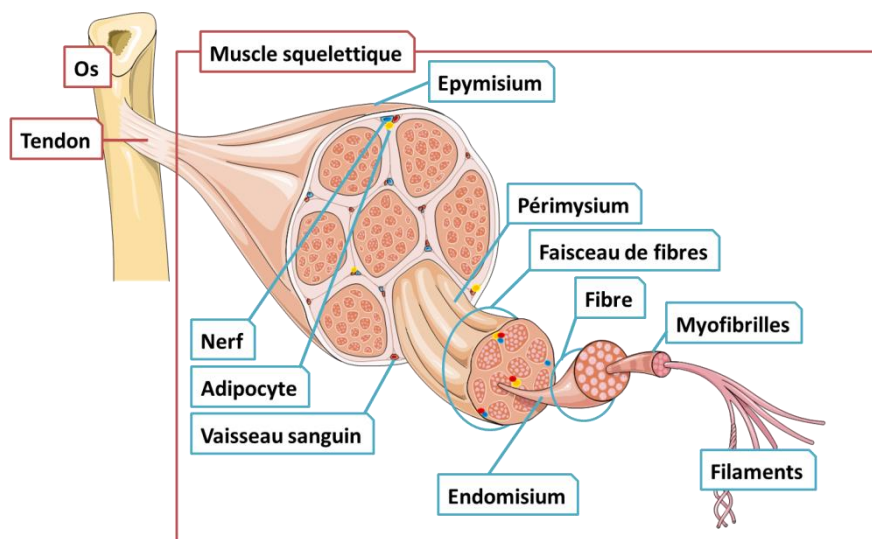


Image : Emma Kemp avec Servier Medical Art (eurostemcell.org).

Figure 1 : Organisation structurale d'un muscle squelettique. Il s'agit d'une coupe du muscle selon les différentes parties qui le compose (de la fibre et ses filaments jusqu'au tissu musculaire). Dans ce sens, le muscle est composé de nombreux faisceaux de fibres entourés par l'épymysium. Des nerfs, des vaisseaux sanguins et du tissu adipeux sont présents entre les faisceaux de fibres. Ces faisceaux sont eux-mêmes composés de nombreuses fibres elles-mêmes composées de nombreuses myofibrilles, elles-mêmes composées de nombreux myofilaments.

Les différents types de fibres du muscle squelettique adulte

Les fibres musculaires sont plurinucléées et sont entourées d'une membrane plasmique (le sarcolemme). Les noyaux sont disposés à la périphérie de la fibre, les mitochondries et autres organites de la cellule sont répartis dans le sarcoplasme. Chaque fibre est composée d'un assemblage d'éléments contractiles : les myofibrilles. Leur analyse microscopique révèle l'alternance tous les 2,5µm d'une zone appelée sarcomère. Ces sarcomères sont composés de filaments fins (constitués essentiellement d'actine) et de filaments épais (constitués de myosine) (Figure 1 et Figure 2). Il existe différents types de fibres musculaires qui sont classés selon leurs propriétés contractiles, métaboliques (Tableau 1) et leurs couleurs (Schiaffino et Reggiani, 2011).

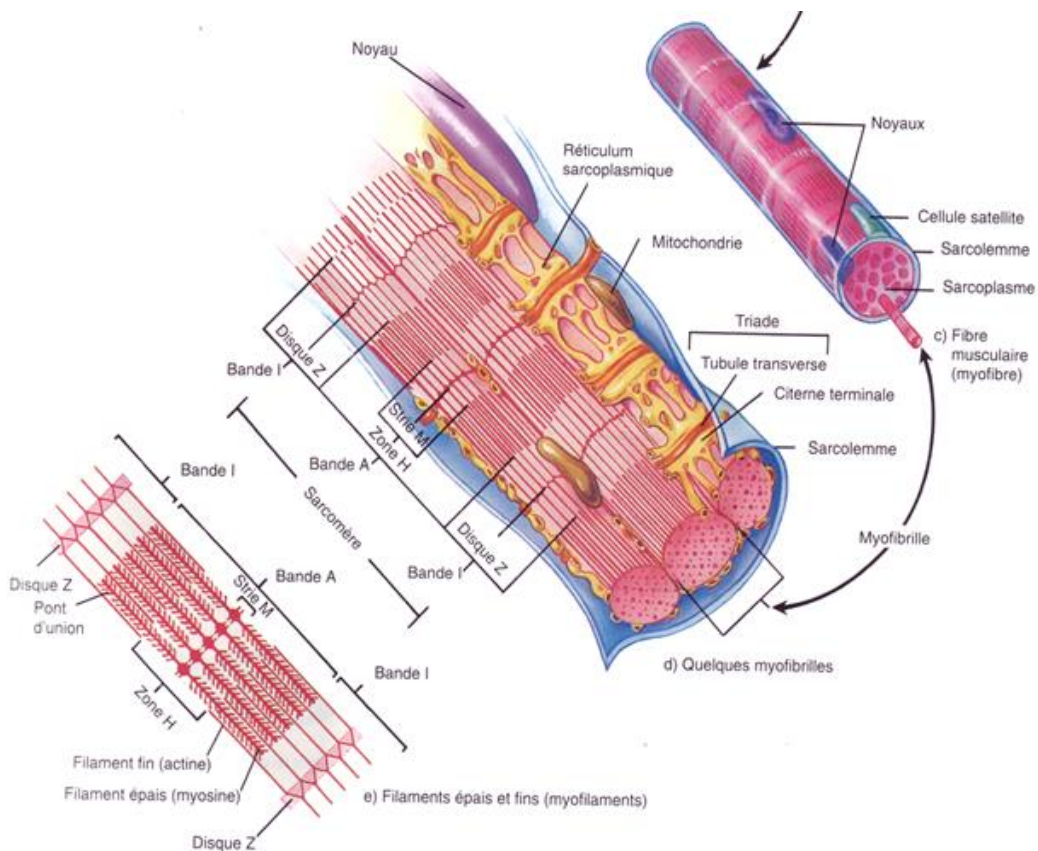


Figure 2 : Organisation structurale d'une fibre musculaire en microscopie électronique (Tortora et Grabowski, 1994). On définit un sarcomère comme étant le segment entre deux disques Z voisins (de l'allemand « zwischen », signifiant « entre »). Dans les coupes longitudinales des muscles en microscopie électronique, les disques Z apparaissent comme une série de lignes foncées. À côté d'un disque Z, on retrouve la bande I (pour isotropique). La bande A (pour anisotropique) vient ensuite. On retrouve une partie plus pâle dans cette région appelée la zone H (de l'allemand « heller », pour « pâle »). Finalement incluse dans cette dernière, la mince strie M (de l'allemand « mittel », signifiant « centre ») se situe à l'intérieur. Les bandes A et I ont été nommées d'après leurs propriétés sous microscope polarisant. Les bandes A et I, ainsi que la ligne Z, sont visibles au microscope électronique.

La classification des fibres

Les propriétés contractiles d’un muscle sont liées à sa force et sa vitesse de contraction, elles sont dépendantes de l’isoforme de chaîne lourde de myosine (MyHC) qu’elles renferment. Les propriétés métaboliques représentent la capacité du muscle à produire de l’énergie pour la contraction musculaire par voie oxydative ou glycolytique. La combinaison des propriétés contractiles et métaboliques permet de distinguer 4 types de fibres dans le muscle adulte ([Schiaffino et Reggiani, 2011](#)) (Tableau 1). Les fibres dites de type I, ou lentes oxydatives, renferment l’isoforme MyHCI, elles utilisent essentiellement les lipides qui sont oxydés pour produire de l’énergie et sont riches en adipocytes et en mitochondries. À l’opposé, les fibres dites de type IIB, ou rapide glycolytique, renferment l’isoforme MyHCIIb et utilisent essentiellement les glucides comme source d’énergie pour la contraction musculaire. De ce fait, elles renferment très peu d’adipocytes et de mitochondries. Les fibres de type IIA, ou rapides oxydo-glycolytiques, renferment l’isoforme MyHCIIa et sont capables d’utiliser à la fois des glucides et des lipides pour la contraction musculaire. Les fibres IIX, qui renferment l’isoforme MyHCIIx, ont la particularité de présenter des propriétés métaboliques intermédiaires entre celles des IIA et des IIB (Tableau 1). Les fibres IIX ont été mises en évidence tout d’abord chez les rongeurs puis leur existence a été démontrée dans d’autres espèces comme le porc ([Lefaucheur, 2010](#)). Chez les espèces comme l’homme, le mouton, le bovin ou le cheval, la MyHCIIb est considérée comme non exprimée ([Picard et al., 2002](#)). Toutefois, chez le bovin [Picard et Cassar-Malek, 2009](#) ont montré que malgré la présence de l’ARNm, la protéine ne se retrouve que chez certains bovins, en particulier chez les races à viande françaises comme la Limousine et la Blonde d’Aquitaine.

Caractéristiques	Type I	Type IIA	Type IIX	Type IIA
<i>Contractile</i>				
Taille et diamètre	Petite	Large	Large	Très large
Vitesse de contraction	+	++	++	+++
Force de contraction	+	++	++	+++
Isoforme MyHC	MyHC I	MyHC IIa	MyHC IIx	MyHC IIb
<i>Métabolique</i>				
Volume mitochondrial	+++	++	++	+
Métabolisme glycolytique	Faible	Élevé	Élevé	Élevé
Métabolisme oxydatif	Élevé	Élevé	Faible	Faible
Adipocytes	Élevé	Élevé	Faible	Faible
Vascularisation	Forte	Forte	Faible	Faible

Tableau 1: Classement des différents types de fibres musculaires en fonction de leurs propriétés métaboliques et contractiles.

Ces différents types de fibres peuvent être distingués sur la base de la séparation des différentes MyHC par électrophorèse à partir d’un broyat musculaire ([Picard et al., 2011](#)). Elles peuvent aussi être classées à partir de coupes de muscles par des techniques histochimiques et immunohistochimiques ([Kammoun et al., 2014](#); [Lefaucheur et al., 1995](#)). L’utilisation d’anticorps spécifiques des isoformes de MyHC par immunohistochimie a l’avantage de permettre la distinction de fibres dites « hybrides » exprimant simultanément plusieurs isoformes de MyHC ([Picard et al.,](#)

2002). Ces fibres hybrides correspondent à des fibres en cours de modification de leurs propriétés métaboliques et contractiles selon le schéma de la [Figure 3](#) en lien avec la plasticité du tissu musculaire.

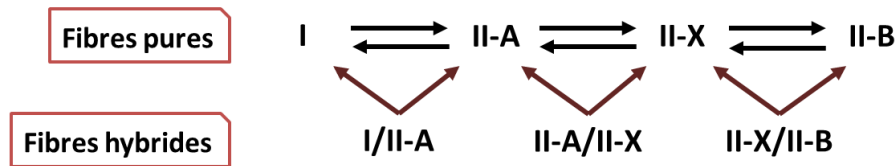


Figure 3 : Plasticité des fibres musculaires. Les fibres pures expriment une seule isoforme de MyHC, les hybrides expriment plusieurs MyHC simultanément

La plasticité des fibres musculaires

La plasticité cellulaire est le phénomène qui permet à des cellules de modifier leurs propriétés structurales et/ou fonctionnelles. Pour le tissu musculaire, la plasticité représente sa capacité à évoluer tout au long de la vie d'un individu en réponse à des facteurs liés à l'animal (âge, sexe) ou à son environnement (alimentation, activité physique). La plasticité du tissu musculaire affecte les fibres qui le composent en leur permettant de changer de type en fonction des besoins de l'individu. La modification des propriétés contractiles des fibres musculaires suit le schéma décrit en [Figure 3](#). Par exemple, une fibre de type I peut devenir une fibre de type IIA en passant par un stade hybride I/IIA. Dans ce cas, le changement démarre par une diminution des isoformes de MyHC-I qui sont progressivement remplacées par des isoformes de MyHC-IIa. On observe une diminution de l'abondance des ARNm MyHC-I et une augmentation des ARNm MyHC-IIa. Ces modifications s'accompagnent de changement des propriétés métaboliques de la fibre. L'exemple de plasticité le plus connu est celui survenant au cours du développement d'un individu. On observe une forte augmentation de la proportion des fibres de type IIX de la naissance à la puberté, les proportions des fibres I et IIa diminuent durant cette même période. Ces proportions restent stables, puis on observe une diminution de la proportion de fibres de type IIX lorsque celles des fibres I et IIA augmentent progressivement avec le vieillissement de l'individu ([Butler-Browne et Bigard, 2006](#)).

1.2 – Les tissus adipeux

Le TA est un tissu spécialisé dans le stockage des lipides, notamment des triglycérides. Chez le bovin, le TA représente 4 à 7% du poids de l'animal à la naissance contre environ 7 à 35% à l'âge adulte en fonction des races ([Bonnet et al., 2010](#); [Gotoh et al., 2009](#)).

Les différents tissus adipeux

Le TA est principalement constitué d'adipocytes, mais aussi d'autres types cellulaires comme les préadipocytes, les macrophages, les cellules endothéliales et les fibroblastes ([Ailhaud et al.,](#)

[1992](#)). La masse du TA dépend du nombre d'adipocytes et de leur volume. Le TA est présent directement sous la peau (sous-cutané) ou au niveau des organes (par exemple le TA périrénal autour des reins), des muscles (TA inter musculaire) ou dans les fibres musculaires (TA intra musculaire). Les adipocytes sont groupés et entourés de tissu conjonctif ([Ailhaud et al., 1992](#)) et d'un réseau de fibres nerveuses plus ou moins important selon la localisation anatomique du TA ([Fliers et al., 2003](#)). Des vaisseaux sanguins sont aussi présents au sein du TA, ils forment un réseau fortement développé pour les TA internes et moins développé dans le TA sous-cutané ([Hausman et al., 2001](#)). Cette forte vascularisation participe à la fonction endocrine des TA et à une communication étroite entre les TA et les autres tissus et organes ([Raschke et Eckel, 2013](#); [Romacho et al., 2014](#)). Dans la majorité des espèces, dont les ruminants, deux types de TA coexistent : le TA brun, principalement présent en fin de vie fœtale et en début de vie post natale chez les mammifères non hibernants ([Casteilla et al., 1989](#); [Smith et al., 2004](#)), à l'exception du porc ([Trayhurn et al., 1989](#)) et durant toute la vie chez les mammifères hibernants exposés au froid ([Cinti, 2005](#)) ; et le TA blanc, essentiellement présent après la naissance chez les mammifères non exposés au froid ([Feve, 2005](#); [Novakofski, 2004](#)). Cependant, le concept couramment admis depuis des décennies selon lequel les TA fœtaux sont exclusivement bruns chez les grands mammifères, a été revu suite à la caractérisation cellulaire et moléculaire des TA périrénal ([Taga et al., 2011](#)) et intermusculaire ([Taga et al., 2012](#)) des fœtus bovins. En effet, ces travaux ont mis en évidence que les TA fœtaux de bovins sont hétérogènes et composés d'un mélange d'adipocytes bruns et blancs. Enfin, plus récemment un troisième type de TA nommé « beige » a été identifié chez les rongeurs: il s'agit d'adipocytes beiges qui émergent dans les tissus adipeux blancs dans certaines conditions comme l'exposition au froid ([Petrovic et al., 2010](#); [Sanchez-Gurmaches et Guertin, 2014](#)). Leur présence a aussi été suggérée chez l'homme ([Pisani et al., 2011](#); [Wu et al., 2012](#)) et le mouton ([Pope et al., 2014](#)).

Les différentes fonctions des adipocytes

Les adipocytes blancs, bruns et beiges n'ont pas les mêmes propriétés structurales et biologiques ([Tableau 2](#)). L'adipocyte blanc mature est sphérique, chez le ruminant son diamètre varie de 25 à 250µm. L'adipocyte blanc possède un noyau (et des mitochondries) repoussé(s) à la périphérie de la cellule par une vacuole lipidique unique et très volumineuse ([Purroy, 1997](#); [Robelin, 1990](#)). L'adipocyte blanc est composé en majeure partie de lipides, avec une faible quantité d'eau, de protéines et de minéraux. L'adipocyte brun a une teneur en lipides et en protéines plus faible que les adipocytes blancs ([Taga et al., 2012](#)) et en conséquence a une teneur en eau plus élevée. Les adipocytes bruns sont moins ronds que les adipocytes blancs et chez les ruminants leur taille varie de 10 à 75µm. Ils sont composés d'un noyau rond situé au centre de la cellule et entouré de plusieurs petites vacuoles lipidiques ([Smith et al., 2006](#)) et de nombreuses mitochondries ([Landis et al., 2002](#)). Enfin, l'adipocyte beige est similaire au brun par sa forme et sa composition, cependant il possède des gouttes de lipides plus grosses que celles de l'adipocyte brun et possède moins de mitochondries ([Nedergaard et Cannon, 2013](#))([Figure 4](#)).

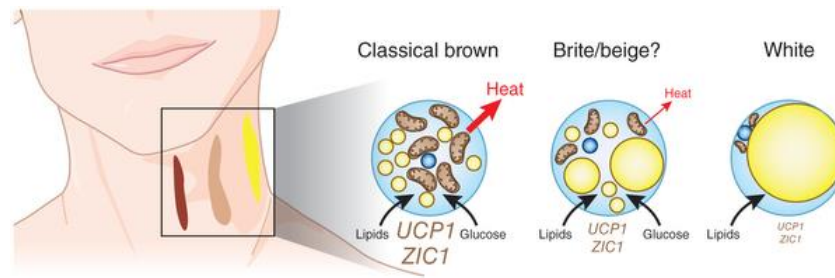


Figure 4 : Différences structurales entre les 3 types d'adipocytes ([Nedergaard et Cannon, 2013](#)).

La fonction principale des adipocytes est de stocker les triglycérides (TG). La quantité de TG stockée dépend d'un équilibre entre la lipogenèse (trois voies de lipogenèse : hydrolyse des acides gras circulants, synthèse *in situ* d'acides gras et estérification (synthèse de TG) des acides gras captés ou synthétisés *in situ*) et la lipolyse (hydrolyse des lipides stockés : sécrétion des acides gras et de glycérol dans le sang). En outre, la spécificité des adipocytes bruns est de produire de la chaleur (fonction de thermogenèse) par oxydation des acides gras mobilisés ([Cannon et Nedergaard, 2004](#)). La lipogenèse a lieu lorsque le bilan énergétique des animaux est positif (surplus d'acide gras) et la lipolyse lorsque le bilan énergétique est négatif ([Bonnet et al., 1998](#); [Bonnet et al., 2004](#); [Chilliard, 1993](#)). Les fonctions des adipocytes beiges sont encore mal connues. Cependant, ils possèderaient une fonction de thermogenèse similaire à celle des adipocytes bruns. Il est à noter qu'une trans-différenciation des adipocytes blancs en beiges (et inversement) existe sous certaines conditions environnementales comme le froid ([Park et al., 2014](#)). Les adipocytes ont aussi une fonction de sécrétion (adipokines et cytokines) et communiquent avec d'autres organes notamment pour le maintien de l'homéostasie énergétique (balance énergétique de l'organisme ([Chilliard et al., 2000](#))).

Caractéristiques	TA Brun	TA Blanc	TA Beige
Morphologie	Un noyau, plusieurs petites gouttes lipidiques et nombreuses mitochondries	Un noyau aplati, une grande goutte lipidique et peu de mitochondries	Un noyau, plusieurs gouttes lipidiques de taille intermédiaire et des mitochondries
Fonction	Production de chaleur	Stockage de l'énergie	Stockage des lipides ou production de chaleur en fonction des besoins
Mitochondrie	+++	+	++
Vascularisation	+++	+	+
Activateurs	Froid FGF21, BMP7, BMP8b, activation par les PPAR et les β -adrénergiques	Apport trop important d'énergie	Froid FGF21, activation par les PPAR
Marqueurs	UCP1, EVA1, PDK4, EBF3, HSPB7	LPL, G3PDH	TBX1, CITED1, SHOX2, TMEM26

Tableau 2 : Différences structurales et fonctionnelles entre les 3 types d'adipocytes (adapté de ([Park et al., 2014](#))).

2 - L'ontogenèse du muscle et du TA

Le développement du muscle et des tissus adipeux résulte de nombreux processus biologiques se déroulant dès le stade embryonnaire, et se poursuivant pendant la croissance foetale et post-natale, on parle de l'ontogenèse des tissus. Elle se déroule en plusieurs étapes : l'embryogenèse (prolifération des cellules souches embryonnaires (CSE)), la détermination (les CSE se différencient en précurseurs musculaires ou adipeux), la prolifération ou hyperplasie des précurseurs musculaires et adipeux, la différenciation et l'hypertrophie des cellules caractérisées par l'acquisition de leurs fonctions biologiques. Chaque tissu possède une chronologie de développement spécifique. L'ontogenèse du muscle précède l'ontogenèse des TA chez les mammifères.

2.1 – Ontogenèse du muscle

Les précurseurs myogéniques ont pour origine une zone de l'embryon appelé myotome ([Christ et Brand-Saberi, 2002](#)). Plus précisément, ces précurseurs sont issus de structures appelées somites. Ce sont des structures sphériques épithéliales formées de part et d'autre du tube neural embryonnaire, nécessaires au développement de toute la structure de l'embryon. Ils donnent naissance à des structures du développement embryonnaire ([Buckingham et al., 2003](#)) : le sclérotome (à l'origine des vertèbres, des côtes et des cartilages axiaux), le dermomiotome (à l'origine des muscles striés squelettiques) et le dermatome (à l'origine des tissus conjonctifs de la peau). La somitogenèse est un processus conservé chez tous les vertébrés, qui a lieu lors des phases précoces embryonnaires à la fin de la gastrulation. L'engagement de la cellule somitique vers un lignage musculaire dépend de l'expression de régulateurs transcriptionnels myogéniques de la famille MyoD ([Weintraub et al., 1991](#)). Les différents régulateurs myogéniques exercent une fonction différente sur la détermination et la différenciation des précurseurs musculaires ([Arnold et Winter, 1998](#)) ; voir paragraphe 3.2). Après la phase de détermination des précurseurs musculaires, le processus de mise en place du tissu musculaire ou myogenèse, se fait en 3 phases : la prolifération des myoblastes, leur fusion en myotubes et la différenciation des myotubes en fibres musculaires ([Figure 5 ; Biressi et al., 2007](#)).

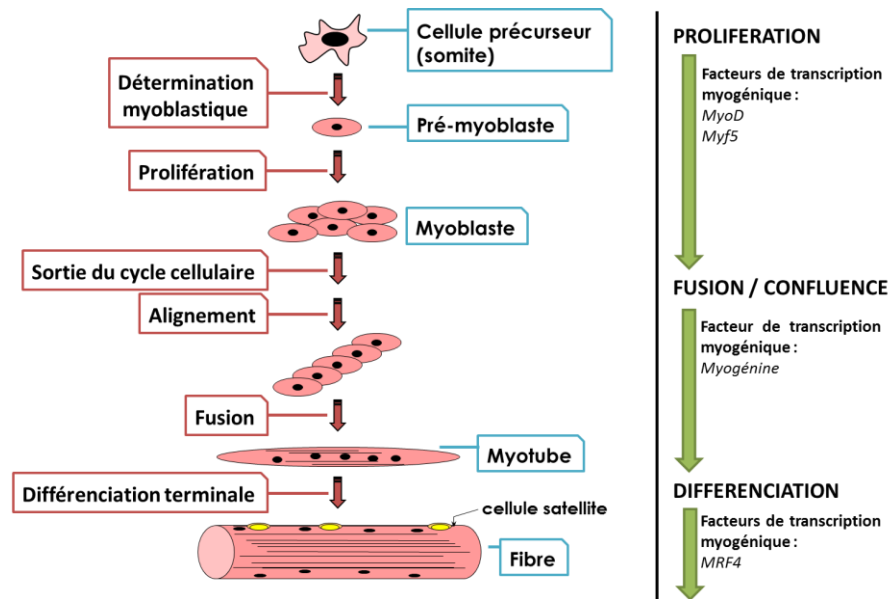


Figure 5 : Principales étapes de la myogenèse.

La myogenèse chez les mammifères

La première phase de la myogenèse, appelée hyperplasie, est caractérisée par une prolifération des myoblastes. La fin de la prolifération des myoblastes se fait sous le contrôle de facteurs de régulation du cycle cellulaire. Les myoblastes sortent du cycle cellulaire et fusionnent en myotubes (Figure 5). Dans toutes les espèces, plusieurs générations successives de myoblastes ont été décrites (Wigmore et Evans, 2002). Les myoblastes primaires (myoblastes embryonnaires) sont principalement à l'origine des fibres lentes chez le fœtus. Les myoblastes secondaires (myoblastes fœtaux) apparaissent en cours de vie fœtale et sont en majorité à l'origine de fibres rapides, mais aussi de fibres lentes pour les muscles mixtes (à fibres lentes et rapides). Une troisième génération de myoblastes a été décrite chez plusieurs espèces comme l'homme (Draeger et al., 1987), le porc (Lefaucheur et al., 1995), le mouton (Wilson et al., 1992) et le bovin (Picard et al., 2002). Cette troisième génération a été identifiée autour de la moitié de la vie fœtale. Le nombre total de fibres (NTF) augmente et se stabilise en fin de vie fœtale ou autour de la naissance selon les espèces et ne sera plus modifié au cours de la vie postnatale (Picard et al., 2010). Chez le bovin comme chez l'homme, le NTF est fixé aux environs des deux tiers de la vie fœtale. La seconde phase de la myogenèse est caractérisée principalement par une augmentation de la taille des fibres (hypertrophie musculaire) et par l'acquisition des propriétés contractiles et métaboliques des fibres. La croissance post-natale du muscle se fait par hypertrophie musculaire et est caractérisée par une modification des propriétés des fibres appelée plasticité.

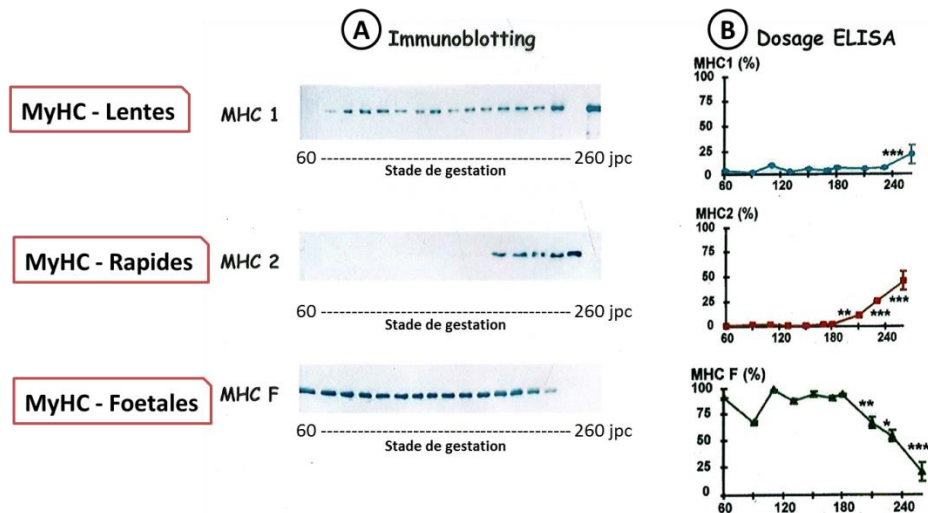
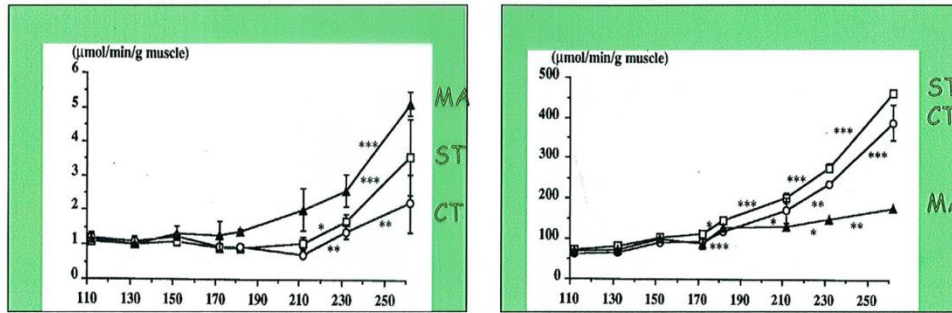


Figure 6 : Evolution des isoformes de chaînes lourdes de myosine (MyHC) au cours de la vie fœtale chez le bovin (Picard *et al.*, 2003) suivie à l'aide d'anticorps spécifiques

La myogenèse chez le bovin

Chez le bovin, les myoblastes primaires sont observés à partir d'environ 30 jpc, les myoblastes secondaires à partir de 90 jpc et les myoblastes tertiaires à partir de 110 jpc. La première génération, représentant environ 30% du NTF chez l'adulte, est à l'origine principalement des fibres de type I. La deuxième génération donne des fibres de type IIX et la troisième génération donne des fibres de type IIA et IIB (Picard *et al.*, 2002). La prolifération des myoblastes primaires a lieu environ de 30 à 110 jpc, celle des myoblastes de seconde et troisième générations jusqu'à 180jpc, âge autour duquel le NTF est fixé chez le bovin (Picard *et al.*, 2002). La différenciation contractile des fibres musculaires est suivie au travers de l'expression des différentes isoformes de MyHC. Les isoformes embryonnaire et fœtale présentes dès le début de la gestation sont progressivement remplacées par les isoformes adultes. Les fibres issues des myotubes primaires n'expriment plus que de la MyHC lente à partir d'environ 180 jpc (Picard *et al.*, 1994; Robelin *et al.*, 1993). Dans les fibres secondaires et tertiaires, la MyHC fœtale disparaît progressivement au cours du troisième tiers de la vie fœtale (Figure 6) et a totalement disparu dans les jours suivant la naissance. Ces résultats démontrent que la différenciation contractile du muscle bovin est très avancée à la naissance comparativement à d'autres espèces telles que les rongeurs chez lesquels l'isoforme fœtale disparaît un mois après la naissance (Schiaffino et Reggiani, 2011). La différenciation métabolique, caractérisée par une augmentation des activités des enzymes oxydatives et glycolytiques, a lieu au cours du dernier tiers de la vie fœtale (Figure 7). La différenciation métabolique des fibres primaires est terminée autour de 180 jpc (Duris *et al.*, 1999). Celle des fibres secondaires et tertiaires se termine autour de la naissance. Ainsi, les muscles du bovin sont distingués en fonction de leurs propriétés contractiles et métaboliques dès la fin de la vie fœtale.



Isocitrate déshydrogénase (ICDH) Lactate déshydrogénase (LDH)

MA : Masseter (oxydatif) - ST : Semitendinosus (mixte glycolytique) - CT : Cutaneus trunci (glycolytique)

Figure 7 : Evolution des activités enzymatiques de la LDH et de l'ICDH au cours de la vie fœtale dans différents types de muscles (Gagniere et al., 1999). La différenciation métabolique des muscles est précoce et a lieu au cours de la vie fœtale. Le métabolisme des fibres est déterminé par les activités enzymatiques des fibres, notamment de l'isocitrate déshydrogénase (ICDH ; métabolisme oxydatif) et de la lactate déshydrogénase (LDH ; métabolisme glycolytique). Ces deux enzymes représentent des voies métaboliques qui produisent de l'ATP, l'énergie nécessaire à la contraction des muscles. Le métabolisme oxydatif est celui des fibres de type I et IIA qui produisent plus d'ATP par molécule de glucose que les fibres rapides à métabolisme glycolytique (type IIX et IIB ; Tableau 1).

La myogenèse chez le bovin a été bien décrite, en particulier grâce à l'analyse protéomique du muscle à différentes périodes de la vie fœtale caractérisant les phases de 60 à 180 jpc, puis de 180 jpc à la naissance (Picard et al., 2010). De nombreuses protéines jouant un rôle dans la prolifération ou la différenciation des myoblastes ont ainsi été identifiées (Figure 8). Ces résultats ont été obtenus sur le muscle *semitendinosus* (ST, muscle de la cuisse) qui est facilement prélevable et individualisé dès 60 jours de vie fœtale.

De 60 à 180 jpc : la prolifération et le début de la différenciation

Entre 60 et 180 jours de vie fœtale, le poids du ST est augmenté d'un facteur 15 et sa longueur d'un facteur 2,5. Cette période constitue une phase majeure pour la croissance du muscle marquée par la modification d'abondance d'un grand nombre de protéines (Chaze et al., 2008). L'analyse protéomique du muscle ST aux stades 60, 110 et 180 jpc a révélé 500 spots protéiques dans un gradient de pH 4-7 (Chaze et al., 2008). Parmi les 173 protéines identifiées par spectrométrie de masse, 157 étaient communes aux 3 stades.

Le stade 60 jpc correspond à la prolifération et fusion des myoblastes de première génération pour donner des myotubes primaires. Ce stade est principalement caractérisé par une forte abondance de protéines impliquées dans le développement embryonnaire, la régulation de l'organisation et la croissance cellulaire. Par exemple, les protéines telles que HNRNPK (Heterogeneous nuclear ribonucleoprotein K), PA2G4 (Proliferation-associated protein 2G4), SEPT11 (Septin 11) et ARHGDI1A (Rho GDP dissociation inhibitor alpha), connues pour jouer un rôle dans la

migration, la prolifération et la régulation du cycle cellulaire (*source UniProt*), sont plus abondantes à 60 jpc.

Le stade 110 jpc correspond plus particulièrement à la prolifération des myoblastes de seconde et troisième génération et leur fusion en myotubes ainsi qu'à la différenciation des premières générations dont la prolifération est terminée. Ainsi, à ce stade, les termes GO les plus retrouvés correspondent principalement à la fusion et à la différenciation cellulaire. Par exemple, les protéines impliquées dans la structure du muscle (MYL2 et MYL6B, respectivement Myosin Light Chain 2 et 6B) et la mise en place d'organites cellulaires (LGALS1 et CLIC1, respectivement Galectin-1 et Chloride intracellular channel protein 1) sont plus abondantes à ce stade ([Chaze et al., 2008](#)). On retrouve aussi DES (Desmine) et LGALS1, des protéines connues comme marqueur de la myogenèse chez la souris ([Poirier et al., 1992](#); [Yablonka-Reuveni et al., 1999](#)).

Le stade 180 jpc correspond à la fin de prolifération des myoblastes et donc à la transition entre prolifération et différenciation des cellules musculaires. En accord avec ceci, les termes GO les plus représentés correspondent à la différenciation cellulaire, à la signalisation cellulaire, la génération de précurseurs métaboliques. À ce stade, des protéines liées à la structure du muscle (MYL4, MYL2, TNNT3 et TNNT1, respectivement des Myosin Light Chain et Troponin T), au métabolisme énergétique (MYL4, AHSA1, PTGR2 et FAM213B, respectivement Myosin Light Chain, Activator of protein ATPase, Prostaglandine reductase et Prostaglandine F synthase) et à la régulation de l'apoptose (DDAH2, ALB, ANXA4 et HSPB1, respectivement Dimethylarginine Dimethylaminohydrolase, Serum albumine, Annexine A4 et Heat shock protein beta-1) sont plus abondantes qu'aux deux autres stades ([Chaze et al., 2008](#)). On observe aussi les Troponines TNNT3 et TNNT1 qui sont liées à l'acquisition des propriétés de contraction du muscle ([Robinson et al., 2007](#)).

Certaines protéines ont des abondances qui évoluent en fonction du stade considéré. C'est notamment le cas de la protéine CLIC4 décrite comme un marqueur de la prolifération des myoblastes primaires ([Ronnov-Jessen et al., 2002](#)). Dans le muscle ST de bovin, cette protéine est surabondante de 60 à 110 jpc et sous-abondante à 180 jpc. CLIC4 est aussi un effecteur de l'apoptose *via* une activation des caspases. Une interaction entre CLIC4 et Park7, impliquée dans la protection contre le stress oxydant a été décrite ([Chaze et al., 2008](#)). La desmine, protéine de structure ([Goldfarb et al., 1998](#)) impliquée dans l'organisation du cytosquelette ([Munoz-Marmol et al., 1998](#)), est abondante très tôt dans la vie fœtale bien avant le facteur de transcription MyoD (Myoblast determination protein). La desmine est surabondante à 110 jpc et pourrait être un marqueur du début de différenciation des myoblastes primaires. Les protéines HNRNPK et HSPB1 (HSP27), connues pour leur rôle anti apoptotique ([Bruneel et al., 2005](#); [Carpenter et al., 2006](#)), sont plus abondantes de 60 jpc à 110 jpc puis leur abondance diminue lorsque le NTF est fixé à partir de 180 jpc. Ces résultats montrent que le mécanisme d'apoptose est nécessaire pour réguler la prolifération des myoblastes et ainsi le nombre total de fibres.

Ainsi, les études de protéomique de Chaze et al. ([Chaze et al., 2008](#)) ont révélé qu'entre 60 et 180 jpc, les protéines les plus abondantes ont un rôle dans la prolifération cellulaire, l'apoptose, l'organisation du cytosquelette. Certaines de ces protéines peuvent être considérées comme des marqueurs de la prolifération des myoblastes et du début de la différenciation.

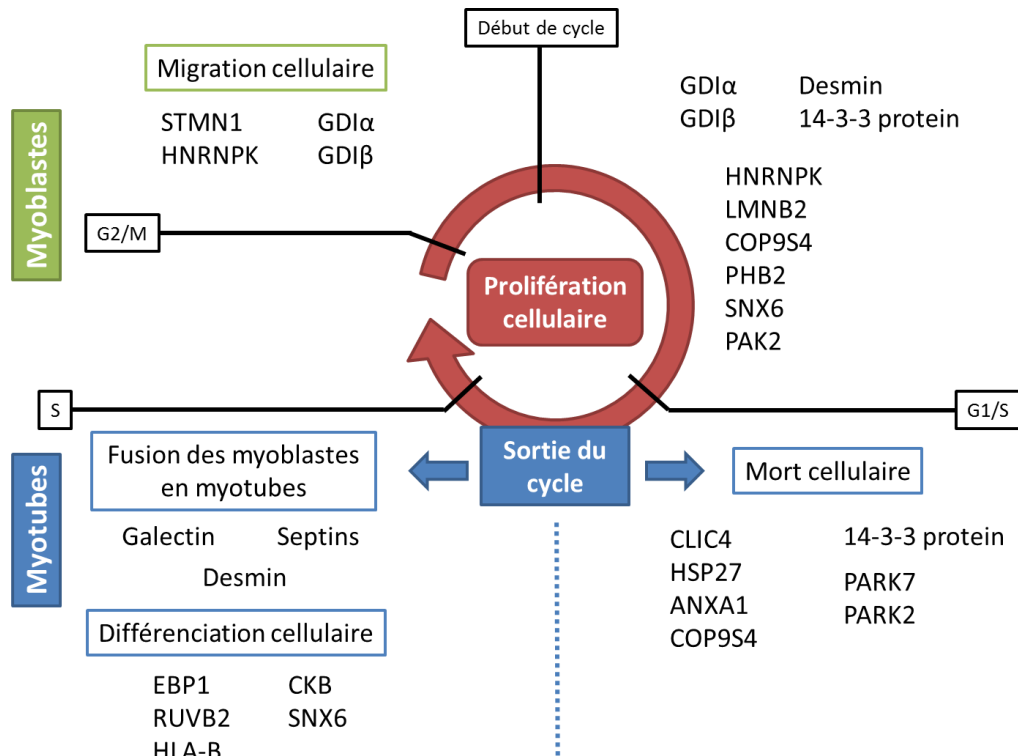


Figure 8 : Synthèse regroupant différentes protéines identifiées ayant un rôle connu ou hypothétique dans la migration cellulaire, la prolifération des myoblastes, la sortie du cycle cellulaire conduisant à la fusion des myoblastes et leur différenciation en myotubes ou à leur mort par apoptose (adapté de (Chaze *et al.*, 2008)). Nous n'avons pas mentionné les protéines impliquées dans la différenciation métabolique et contractile des fibres musculaires. Les étapes de la myogenèse telles que la migration des précurseurs et leur prolifération se déroulent en même temps. Il en est de même pour les étapes de fusion et de différenciation des myotubes des différentes générations.

De 180 à 260 jpc : la différenciation et l'acquisition des fonctions biologiques

Les données histologiques (Picard *et al.*, 2003) et protéomiques (Chaze *et al.*, 2009) démontrent que le dernier tiers de vie fœtale (de 180 jpc à la naissance) est caractérisé par la maturation des fibres musculaires, et l'acquisition de leurs propriétés contractiles et métaboliques.

Parmi les 500 spots protéiques révélés par Chaze *et al.* (Chaze *et al.*, 2008; Chaze *et al.*, 2009) 169 étaient communs aux 3 stades 180, 210 et 260 jpc. Les protéines les plus abondantes à 180 jpc sont impliquées dans le contrôle du cycle cellulaire (PSMB7 et PSME1, respectivement Proteasome subunit beta type-7 et Proteasome activator complex subunit 1), la croissance ou la mort cellulaire (ALB, DDAH2, ARHGDI, HSPB1, ANXA4, ENO1 et PHB, auxquelles s'ajoute l'enolase 1 et la Prohibitine). Les protéines PHB et WARS (Tryptophan-tRNA ligase), dont l'abondance diminue de 180 à 260 jpc, sont impliquées dans la régulation négative de la prolifération et de la croissance cellulaire (Dai *et al.*, 2008; Rubin *et al.*, 1991; Wang *et al.*, 2002). D'autres protéines sont impliquées dans l'acquisition des fonctions biologiques du muscle comme TNNT3 pour la fonction contractile. Les protéines PNPLA1, FAHD2A et ALDH9A1 (respectivement Patatin-like phospholipase protein 1, Fumarylacetoacetate hydrolase et 4-triméthylaminobutyraldehyde dehydrogenase) pour les

métabolismes oxydatifs et glycolytiques ([Kurys et al., 1989](#); [Wilson et al., 2006](#)) voient leur abondance augmenter à partir de 180 jpc ([Chaze et al., 2009](#)).

Le stade de 210 jpc révèle des protéines impliquées dans la constitution des structures de filaments d'actine et de myosine des fibres (FSCN1, CFL2, WDR1 et ACTA1, respectivement Fascine, Cofiline-2, WD repeat-containing protein 1 et alpha skeletal muscle Actin). ACTA1 est une protéine qui est aussi impliquée dans l'adaptation des fibres ([Seo et al., 2006](#)) et la contraction musculaire ([Nowak et al., 1999](#)).

Le dernier âge de la myogenèse étudié, 260 jpc, est composé de protéines impliquées dans l'acquisition des fonctions du muscle. On observe des protéines impliquées dans le métabolisme glycolytique (ENO3 et TPI1, Triosephosphate isomerase), dans le métabolisme oxydatif (PRDX2 et IDH3A ou ICDH, respectivement Peroxiredoxine-2 et Isocitrate dehydrogenase [NAD] subunit alpha) et le mécanisme de contraction musculaire (TNNT3 et MYBPH, Myosin-binding protein H). On observe différentes isoformes de MyLC (1 et 2), de Tubuline et d'ALDH (Aldehyde dehydrogenase) ([Chaze et al., 2009](#)).

À chacun des 3 âges, les protéines impliquées dans la synthèse des protéines sont abondantes, ce qui est en lien avec l'hypertrophie des fibres musculaires observée durant cette période. Le dernier tiers de la vie fœtale bovine est caractérisé par l'acquisition des fonctionnalités métaboliques et contractiles du muscle, illustrée principalement par des modifications dans les isoformes de protéines existantes. Ceci est en accord avec l'augmentation de l'activité des enzymes du métabolisme oxydatif et glycolytique décrite au cours de cette période ([Picard et al., 2002](#)). Par exemple, l'abondance de l'isoforme alpha de l'enolase (ENO1), détectée dès le début de la vie fœtale, diminue à partir de 180 jpc. Dans le même temps, l'abondance de l'isoforme bêta (ENO3) augmente, indiquant que l'isoforme alpha est progressivement remplacée par l'isoforme bêta présente dans le muscle adulte ([Picard et al., 2010](#)). On retrouve aussi des protéines de la famille des Aldéhydes déshydrogénases (ALDH) qui sont impliquées dans le métabolisme des acides aminés, des lipides, des carbohydrate et des vitamines ([Vasiliou et al., 2000](#)). Comme décrit pour les isoformes de MyHC, d'autres protéines contractiles comme MYLC et des TNNT subissent un remplacement d'isoformes développementales par des isoformes adultes entre 180 et 260 jpc ([Bouley et al., 2004](#); [Chaze et al., 2009](#); [Muroya et al., 2003](#)).

Une analyse transcriptomique confirme l'importance des stades 180 et 260 jpc, auxquels l'expression d'un grand nombre de gènes est modifiée ([Sudre et al., 2003](#)). En particulier, l'expression de Trip15 (thyroid receptor interacting protein 15) impliquée dans l'activation et la stabilisation du facteur c-Jun, régulateur de la prolifération, est la plus faible à 260 jpc. Il en est de même par exemple, pour l'activine A, membre de la famille des TGF- β et pour la nébuline. Cette protéique sarcomérique est impliquée dans la régulation de l'organisation des filaments d'actine en lien avec la tropomoduline. La diminution de son expression en fin de vie fœtale est cohérente avec l'organisation structurale des fibres musculaires qui a lieu plutôt durant les deux premiers tiers de vie fœtale ([Picard et al., 2002](#)). Au contraire l'expression de plusieurs protéines du métabolisme oxydatif est augmentée durant le dernier trimestre en accord avec la différenciation métabolique et contractile décrite pour cette période ([Sudre et al., 2003](#)).

Ces étapes de la myogenèse semblent communes aux différents muscles ou races bovines, avec cependant des différences dans la chronologie ou l'équilibre prolifération/différenciation ([Picard et al., 2010](#); [Picard et al., 2002](#)). Par exemple, les bovins de type « culard » ont une augmentation des masses musculaires qui a pour origine une plus grande prolifération cellulaire des myoblastes de seconde génération ([Deveaux et al., 2001](#); [Picard et al., 1998](#)). Cette prolifération supérieure s'accompagne d'une différenciation contractile et métabolique plus tardive. Cette musculature plus élevée présente un métabolisme plus oxydatif que glycolytique par rapport à des bovins non culards ([Gagniere et al., 1997](#)). Les événements moléculaires à l'origine de différences dans l'équilibre prolifération/différenciation ont été décrits en particulier par Rodriguez *et al.* ([Rodriguez et al., 2014](#)).

La régulation de la myogenèse

La régulation de la myogenèse la plus décrite dans la bibliographie est celle par les facteurs de transcription myogéniques. Il existe 4 facteurs de transcription myogéniques : MyoD, myogénine, Myf5 et Myf6 (aussi appelé MRF4) ([Perry et Rudnick, 2000](#)). Myf5 et MyoD sont exprimés en phase de prolifération des myoblastes, myogénine lorsque les myoblastes sortent du cycle cellulaire (fin de prolifération) et MRF4 est exprimé lors de la maturation des myotubes. L'inactivation du facteur MyoD ou Myf5 n'affecte pas la formation du muscle squelettique. Cependant, dans le premier cas l'expression de Myf5 est accrue chez la souris ([Rudnicki et al., 1992](#)) et dans l'autre cas la mise en place du myotome est retardée jusqu'à l'expression de MyoD ([Braun et al., 1994](#)). Des souris doubles mutantes (Myf5⁻ et MyoD⁻) présentent une absence de myoblastes et de fibres musculaires, ce qui démontre qu'au moins un des deux facteurs est nécessaire pour la mise en place du tissu musculaire ([Rudnicki et al., 1993](#)). L'inactivation de myogénine conduit à l'obtention d'un grand nombre de myoblastes non différenciés et ainsi à une absence de fibres musculaires ([Hasty et al., 1993](#)). Une corrélation entre l'expression de myogénine et le NTF a déjà été mise en évidence ([Vivian et al., 1999](#)) et montre l'importance de ce facteur pour le développement du tissu musculaire. L'inactivation de MRF4 n'influence pas le développement du tissu musculaire, car elle est compensée par une augmentation de l'expression de myogénine ([Patapoutian et al., 1995](#); [Zhang et al., 1995](#)). L'expression tardive de MRF4 montre son implication dans la maturation des fibres. Cependant, MRF4 pourrait intervenir avant MyoD et agir sur la détermination du lignage musculaire ([Kassar-Duchossoy et al., 2004](#)). Enfin, des souris doubles mutantes (myogénine⁻ et MRF4⁻) présentent des myoblastes et des myotubes tout aussi efficaces que des témoins et montrent que myogénine et MRF4 ne sont pas indispensables à la différenciation des myoblastes ([Rawls et al., 1998](#)).

D'autres facteurs interviennent dans la régulation de la myogenèse *via* le contrôle du cycle cellulaire ou de la différenciation musculaire. On peut citer les facteurs de croissance tels que FGF (Fibroblast Growth Factor), TGF- β (Transforming Growth Factor), BMP (Bone Morphometric Protein) ou encore les IGF (Insulin Growth Factor). Différentes hormones comme l'insuline, l'hormone de croissance (growth hormone ; GH) ou les hormones thyroïdiennes, sont également impliquées dans la régulation de la myogenèse, mais nous ne détaillerons pas cette partie ([Florini et al., 1991](#)).

2.2 – Ontogenèse des tissus adipeux

Tout comme le muscle, la croissance du TA débute très tôt durant la vie de l'animal et survient essentiellement après la naissance. Cet accroissement de masse des TA résulte de l'augmentation du nombre et du volume des adipocytes. Les contributions relatives de l'hyperplasie et de l'hypertrophie à la croissance du TA varient selon l'âge, les localisations anatomiques et les espèces ([Cianzio et al., 1985](#); [Vernon, 1986](#)). Chez les espèces qui ont une adiposité forte (chez les humains, le poids des TA représente en moyenne 15% du poids corporel chez le nouveau-né humain) ou moyenne (4 % du poids corporel chez les ruminants) à la naissance ([Figure 6](#)), la croissance des TA résulte d'une forte hyperplasie prénatale et d'une forte hypertrophie post-natale. Ainsi, contrairement aux cellules musculaires le nombre d'adipocytes n'est jamais fixé. Inversement chez les espèces ayant une faible adiposité à la naissance (2% souris du poids corporel chez les rongeurs), la croissance par hyperplasie et hypertrophie a lieu essentiellement après la naissance.

Les connaissances sur l'origine embryonnaire des adipocytes restent fragmentaires. Les progéniteurs d'adipocytes auraient en effet plusieurs origines embryonnaires en fonction de la localisation anatomique et du type de TA. Après la phase de développement embryonnaire, le processus de mise en place du TA, ou adipogenèse, a lieu en deux étapes : la détermination des cellules souches mésenchymateuses (CSM) en adipoblastes et la prolifération et différenciation en préadipocytes. Il s'agit de la différenciation précoce. Enfin, les préadipocytes se différencient en adipocytes matures; il s'agit de la différenciation terminale.

L'adipogenèse chez les mammifères

La première étape de l'adipogenèse est la détermination d'une CSM multipotente en adipoblaste et sa différenciation en préadipocyte. Les processus de détermination d'une CSM vers le lignage adipogénique restent à ce jour mal connus, cela est dû en partie à l'absence de marqueurs moléculaires permettant de caractériser les différents intermédiaires cellulaires de l'adipogenèse ([Gregoire, 2001](#)) et aussi à l'origine multiple des adipocytes ([Figure 9](#)). Ainsi, certains adipocytes bruns ou beiges de la partie haute du corps auraient une origine embryonnaire similaire à celle des cellules myogéniques (voir paragraphe 2.1) : le dermomyotome ([Park et al., 2014](#); [Sanchez-Gurmaches et Guertin, 2014](#)). Les adipocytes blancs des TA céphaliques localisés dans la partie haute du corps auraient pour origine des cellules embryonnaires issues de la crête neurale ([Billon et al., 2007](#)) d'origine ectodermique, alors que l'origine embryonnaire des adipocytes blancs/beiges des TA viscéraux reste à déterminer, mais serait peut-être le mésoderme latéral ([Bonnet et al., 2015](#)). Les cellules issues de la détermination sont des adipoblastes qui ont une forte capacité de prolifération. Les adipoblastes prolifèrent jusqu'à confluence et s'arrêtent en phase G1/S du cycle cellulaire. L'arrêt de la croissance, et non la confluence, semble déterminant pour leur différenciation en préadipocytes puisque des précurseurs adipocytaires en culture primaire se différencient (en passant par une expansion clonale) malgré l'absence de confluence et de contact cellulaire ([Boone et al., 2000](#); [Gregoire et al., 1998](#)). À ce stade, les préadipocytes expriment les marqueurs précoces de l'adipoconversion, tel qu'une protéine de la lipogenèse la lipoprotéine lipase ou des facteurs de transcription comme les C/EBP β , C/EBP δ et le PPAR γ ([Figure 9](#)). Les adipocytes n'ayant pas la capacité de proliférer, ce sont les précurseurs adipeux, les adipoblastes et les pré-adipocytes, qui prolifèrent. Ces étapes de détermination et de prolifération ainsi que leur régulation restent peu

décrites quelles que soient les espèces ([Sanchez-Gurmaches et Guertin, 2014](#)). Quelques protéines ou familles de protéines, ont été identifiées comme intervenant dans la détermination des CSM vers le lignage adipocytaire. Des protéines de la famille des BMP, notamment BMP2 et BMP4 interviennent dans la détermination des progéniteurs adipeux vers le lignage des adipocytes blancs, et BMP7 dans celle des adipocytes bruns et beiges ([Bonnet et al., 2015](#); [Tseng et al., 2008](#)). Chez la souris, les progéniteurs ayant exprimé WT1 (un marqueur du mésoderme latéral) sont à l'origine de 30 à 80% des adipocytes viscéraux ([Bonnet et al., 2010](#); [Chau et al., 2014](#)), mais ne sont pas à l'origine des adipocytes sous-cutanés. Quelques rares protéines commencent à être identifiées pour leur rôle dans la régulation de la prolifération des progéniteurs adipeux. C'est le cas par exemple des Activines qui sont des protéines sécrétées qui appartiennent à la superfamille TGF- β . Les Activines A, B et AB sont fortement abondantes dans les TA humains ([Dani, 2013](#)) et l'activine A a été identifiée comme un activateur de la prolifération et un inhibiteur de la différenciation des progéniteurs adipeux ([Zaragosi et al., 2010](#)). Le facteur de croissance FGF10 est aussi un activateur de la prolifération des adipocytes ([Itoh et Ohta, 2014](#)).

La deuxième phase de l'adipogenèse est la différenciation des préadipocytes en adipocytes matures. Une connaissance assez précise des mécanismes qui contrôlent cette étape a été obtenue par des études *in vitro* sur des lignées cellulaires de préadipocytes blancs et de préadipocytes bruns de rongeurs ([Boone et al., 2000](#); [Gregoire et al., 1998](#); [Lefterova et Lazar, 2009](#); [Seale et al., 2009](#)). Les événements moléculaires correspondent à des changements chronologiques dans l'expression de nombreux gènes sous le contrôle de facteurs de transcription. L'expression de trois familles de facteurs de transcription est induite en cascade: PPAR (peroxysome proliferator-activated receptor), SREBP (sterol regulatory element binding protein) et C/EBP (CCAAT/Enhancer Binding Protein). Par ordre chronologique, il y a d'abord l'expression transitoire de PPAR δ , puis celle de SREBP1/ADD1 (sterol regulatory element binding protein-1c) et enfin celles de C/EBP β et δ . L'expression de ces trois derniers facteurs induit celle de PPAR γ 2 ([Boone et al., 2000](#)). La diminution de l'expression de C/EBP β et δ est concomitante à l'augmentation de l'expression de C/EBP α ([Boone et al., 2000](#)). L'expression simultanée de PPAR γ 2 et de C/EBP α induit alors l'expression des gènes caractéristiques du fonctionnement des adipocytes matures. PPAR γ a un rôle central dans le contrôle de la différenciation terminale en régulant l'expression de nombreux gènes spécifiant des enzymes de la lipogenèse. À ce stade les adipocytes acquièrent enfin les marqueurs tardifs de la différenciation des adipocytes blancs, tels que les protéines impliquées dans le métabolisme des AG et des TG. La différenciation se poursuit par l'acquisition du phénotype très tardif correspondant à la sécrétion des nombreuses adipokines ([Taga, 2011](#)). Les mécanismes moléculaires semblent largement communs à toutes les espèces, bien que la cascade d'expression des régulateurs transcriptionnels clés puisse varier ([Ding et al., 1999](#)). En effet, PPAR γ a un rôle central dans le contrôle de la différenciation terminale des préadipocytes isolés de la fraction stroma vasculaire du TA adulte chez les bovins ([Ohyama et al., 1998](#); [Torii et al., 1998](#)) et du TA foetal chez les ovins ([Soret et al., 1999](#)). Cependant, l'expression de C-EBP α n'a pas été détectée ([Taniguchi et al., 2008](#)) ou n'intervient pas dans le processus adipogénique ([Ohsaki et al., 2007](#)) chez les bovins.

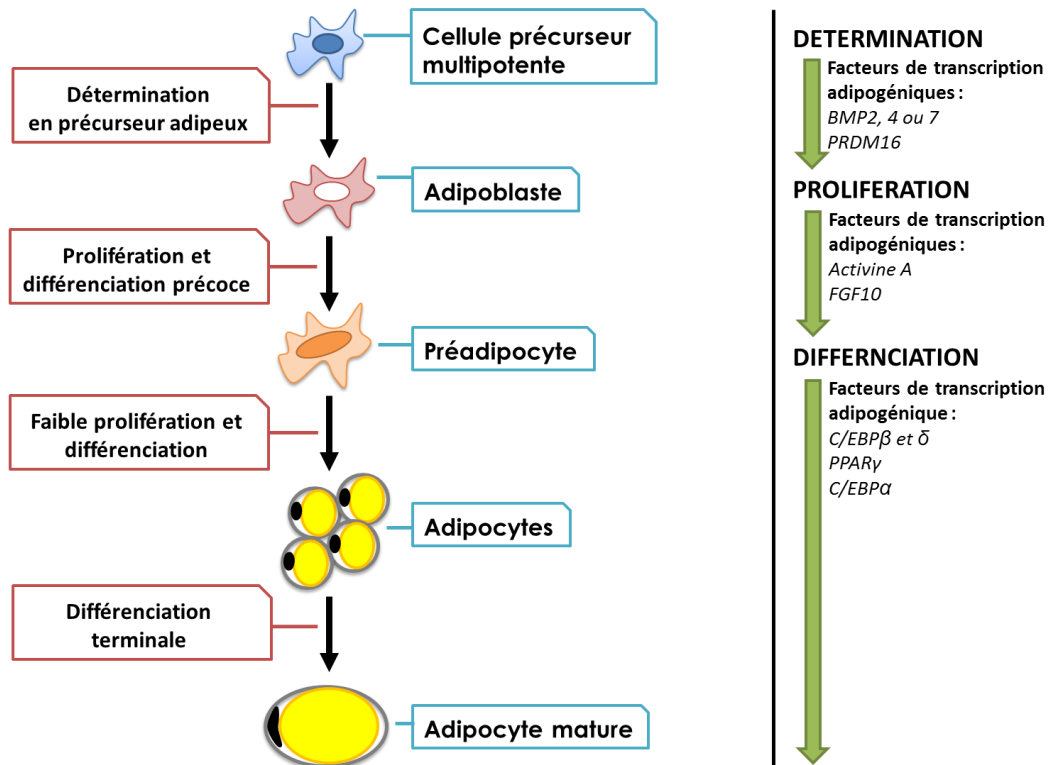


Figure 9 : Principales étapes de l'adipogenèse.

Bien que les adipocytes proviennent de différentes origines cellulaires, les futures différenciations vers un lignage adipocytaire (blanc ou brun) sont régulées par des cascades transcriptionnelles communes qui impliquent principalement les facteurs de transcription des familles PPAR et C/EBP ([Rosen et MacDougald, 2006](#); [Tang et Lane, 2012](#)). Dans certains tissus comme la moelle osseuse, la mélatonine inhibe l'adipogenèse en désengageant les précurseurs du lignage adipocytaire vers le lignage ostéoblastique ([Murdolo et al., 2013](#); [Park et al., 2011](#)). Des études récentes montrent que des protéines, comme la tectorigénine et la motiline ([Li et al., 2015](#); [Miegueu et al., 2011](#)), régulent la sécrétion des adipokines et impactent la différenciation adipocytaire.

Le processus d'adipogenèse s'accompagne de modifications morphologiques des cellules adipeuses puisqu'une cellule fusiforme multiplie son volume et s'arrondit. En particulier la différenciation terminale est marquée par une augmentation du nombre et du volume des vacuoles lipidiques et leur fusion finale en une seule et grosse vacuole pour les adipocytes blancs. L'un des constituants majeur de la membrane de la vacuole est l'adipophiline appelée aussi adipose differentiation related protein (ADRP) qui constitue donc un marqueur précoce de la différenciation adipocytaire ([Gregoire et al., 1998](#)). En outre, ces modifications cellulaires s'accompagnent de variations dans l'abondance, la sécrétion et l'organisation de nombreux composants de la matrice extracellulaire (MEC) ([Lilla et al., 2002](#)).

L'adipogenèse chez le bovin

Chez le bovin, le TA apparaît durant la vie fœtale et son apparition varie selon les sites anatomiques (Robelin, 1990). Les TA périrénal, intramusculaire et sous-cutané apparaissent chez le fœtus respectivement à 80, 110 et 180 jpc (soit 27%, 40% et 66% de la gestation) (Vernon, 1986). Le poids du TA périrénal augmente d'environ 40 fois entre 110 et 260 jpc chez des fœtus bovins Charolais ou Blonde d'Aquitaine. Cette augmentation de poids du TA périrénal résulte d'une augmentation importante du nombre d'adipocytes jusqu'à la fin de la vie fœtale, mais aussi d'une augmentation du volume des adipocytes surtout à partir de 210 jpc (Figure 10 (Taga, 2011)). La synthèse des données disponibles suggère deux périodes importantes pour la croissance fœtale des TA bovins qui correspondraient à la prolifération et à la différenciation précoce des progéniteurs adipeux puis à leur différenciation tardive. Ces données concernent des activités d'enzymes lipogéniques et des expressions de gènes spécifiques aux fonctions des adipocytes blancs et bruns, ainsi qu'une caractérisation du protéome du TA périrénal à 110, 180, 210 et 260 jpc (Taga et al., 2012)

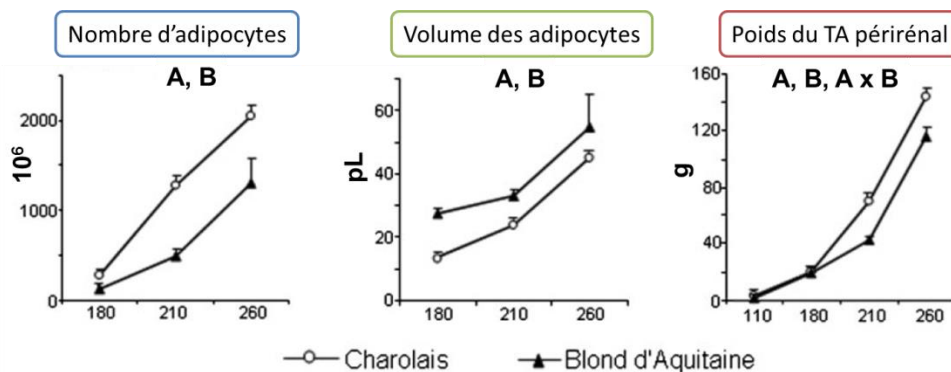


Figure 10 : Nombre total d'adipocytes périrénaux, volume de ces adipocytes (pL) et poids du TA périrénal (g) à 110 (pour le poids seulement), 180, 210 et 260 jpc pour les fœtus des races Charolaise et Blonde d'Aquitaine (adapté de (Taga, 2011)). Les résultats sont des moyennes et des écarts types, A, B, A x B : effets significatifs de l'âge et de la race ($P < 0.001$) et des interactions entre l'âge et la race (à $P < 0.05$).

De 110 à 180 jpc : prolifération et différenciation précoce des progéniteurs adipeux par une régulation du cycle cellulaire et de l'apoptose

Les données disponibles à 180 jpc indiquent que le nombre total d'adipocytes dans le TA périrénal est d'environ 250 millions chez les fœtus bovins de races Charolaise et Blonde d'Aquitaine. Les activités lipogéniques pendant la vie fœtale sont faibles, et elles sont plus faibles à 180 qu'à 210 jpc (Taga et al., 2011; Taga et al., 2012), ce qui explique le petit volume moyen (20 pL, le seuil pour compter un adipocyte étant à 8 pL) des adipocytes à 180 jpc. Ces données suggèrent qu'avant 180 jpc la croissance du TA résulte surtout d'une prolifération des progéniteurs adipeux et de leur différenciation en adipocytes. L'analyse du protéome aux 4 âges fœtaux a indiqué que les protéines dont l'abondance est élevée à 110 et 180 jpc, donc aux âges d'une forte augmentation du nombre d'adipocytes, sont des protéines impliquées dans le contrôle du cycle cellulaire (YWHAE, YWHAG, PSME2, PSMA3, TUBB et TMOD3, qui sont respectivement les protéines 14-3-3 protein epsilon et

gamma, Proteasome activator complex subunit 2, Proteasome subunit alpha type-3, Pyr 60 kDa heat shock protein, mitochondrialvate dehydrogenase E1 component subunit beta et Tropomodulin-3) et de l'apoptose (PSME2, PSMA3, YWHAE, YWHAG, YWHAB, CLIC4, LMNB1, VIM, EIF5A, VCP, PRDX2, HSPB1 et HSPD1, qui sont respectivement 14-3-3 protein beta, Chloride intracellular channel protein 4, Lamin-B1, Vimentin, Eukaryotic translation initiation factor 5A, Transitional endoplasmic reticulum ATPase, Peroxiredoxin-2, Heat shock protein beta-1 (Hsp27) et Heat shock protein &D (Hsp 60)). En effet, les protéines 14-3-3 identifiées sont des régulateurs connus du cycle cellulaire *via* les complexes protéines kinases dépendantes des cyclines et cyclines (cdk/cycline) et des protéines activatrices de l'apoptose ([Mhaweck, 2005](#)). En effet, la fixation de YWHAE ou YWHAG avec Cdc25 ou Wee1 induit un arrêt du cycle cellulaire. Inversement la liaison de YWHAE, YWHAG, YWHAB sur p27 induit une activation du cycle cellulaire ([Mhaweck, 2005](#)). Les protéines YWHAE, YWHAG, YWHAB ont une abondance faible à 110 jpc puis qui augmente entre 180 et 210 jpc. L'hypothèse proposée par Taga et al. ([Taga et al., 2012](#)) est que la faible abondance de ces protéines à 110 jpc est propice à la progression du cycle cellulaire et à la prolifération cellulaire, alors que leur abondance élevée à 180 ou 210 jpc doit réguler l'équilibre entre progression du cycle cellulaire et arrêt du cycle cellulaire, en fonction de la présence des protéines cibles des 14-3-3 et de la phase du cycle cellulaire. Ces mécanismes de progression et d'arrêt du cycle cellulaire sont connus pour impacter la prolifération des précurseurs adipeux ([Gregoire et al., 1998](#)) (voir paragraphe 3). En outre, d'autres protéines identifiées pourraient participer à la régulation de la prolifération des progéniteurs adipeux en raison de leur rôle connu comme régulateur positif (EIF5A ; ([Sun et al., 2010](#))) et négatif de l'apoptose (PRDX2, HSPB1 et HSPD1 ; ([Corton et al., 2008](#); [DeLany et al., 2005](#); [Joo et al., 2010](#))). L'expression du gène HSPD1 a été observée faible en début de prolifération et élevée lors de la confluence de progéniteurs adipeux humains ([Scheideler et al., 2008](#)). Aussi, HSPD1 pourrait être un régulateur négatif de la prolifération des précurseurs adipeux *via* une régulation positive de l'apoptose ([Chandra et al., 2007](#)). Ainsi, ces protéines participeraient à l'augmentation du nombre d'adipocytes en agissant sur l'équilibre progression du cycle cellulaire/apoptose.

D'autres protéines identifiées à ces âges fœtaux chez le bovin participeraient à la régulation du nombre d'adipocytes en agissant sur d'autres événements cellulaires ou voies de signalisation, par exemple, TCP1 (T-complex protein 1 subunit alpha) dont la surabondance à 180 jpc est simultanée à l'augmentation du nombre d'adipocytes. Or, TCP1 a été décrite comme surabondante en début de prolifération par rapport à des progéniteurs non prolifératifs humains ([Scheideler et al., 2008](#)). TCP1 est connue pour jouer un rôle dans l'assemblage des protéines, dans la formation de l'ATP et aussi dans le repliement de l'actine et de la tubuline. PTGR2, et FKBP4 sont connues pour moduler la régulation de l'adipogénèse par leur action inhibitrice sur PPAR γ ([Garin-Shkolnik et al., 2014](#); [Yu et al., 2013](#)). Enfin, cette période de 110 à 180 jpc est aussi caractérisée par des variations dans l'abondance de protéines du cytosquelette ou impliquées dans la production de filament d'actine. Ces fonctions reflètent probablement les changements de structure cellulaire qui accompagnent les différentes phases de l'adipogénèse et sont indispensables à la différenciation tardive des progéniteurs adipeux ([Figure 11](#) ; ([Gregoire et al., 1998](#))).

De 180 à 260 jpc : augmentation du volume des adipocytes par un contrôle de la différenciation et de l'hypertrophie des cellules adipeuses

Entre 180 et 260 jpc l'augmentation du poids du TA périrénal résulte toujours d'une augmentation du nombre d'adipocytes, mais aussi d'une augmentation de leur volume (Figure 10). Cette phase de différenciation tardive et d'acquisition du phénotype « adipocyte » nécessite l'expression de gènes et de protéines liés aux fonctions spécifiques des adipocytes bruns et blancs et donc aux métabolismes énergétiques et au stockage des lipides (Bonnet *et al.*, 2010; Gregoire *et al.*, 1998). Ainsi, les activités lipogéniques sont les plus élevées à 210 jpc (Taga *et al.*, 2011; Taga *et al.*, 2012) et précèdent donc l'augmentation du volume des adipocytes. Simultanément, la caractérisation du protéome adipeux a révélé qu'entre 180 et 260 jpc, une augmentation de l'abondance des protéines impliquées dans le métabolisme glucidique (TPI1, PGAM1, OGDH, ENO1, FBP1, PDHB, ALDH9A1, LDHB, DLD, ALDH2 et MDH1, qui sont respectivement Triosephosphate isomerase, Phosphoglycerate mutase 1, 2-oxoglutarate dehydrogenase, Alpha-enolase, Fructose-1,6-bisphosphatase 1, Pyruvate dehydrogenase E1 component subunit beta, 4-trimethylaminobutyraldehyde dehydrogenase, L-lactate dehydrogenase B chain, Dihydrolipoyl dehydrogenase, Aldehyde dehydrogenase et Malate dehydrogenase) et la β -oxydation (ACADVL, PCCB, ECHS1, ECH1 et ADIPOQ, respectivement Very long-chain specific acyl-CoA dehydrogenase, Propionyl-CoA carboxylase beta chain, Enoyl-CoA hydratase, Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase et Adiponectin). Des protéines liées au métabolisme et au stockage des lipides ont aussi été identifiées (ECHS1, GPD1, PCCB, ME1, GPD2, ACADVL, THEM4, ACAT2, PC, PRDX6 et FAPB4, respectivement Glycerol-3-phosphate dehydrogenase, Propionyl-CoA carboxylase beta chain, NADP-dependent malic enzyme, Glycerol-3-phosphate dehydrogenase, Very long-chain specific acyl-CoA dehydrogenase, Acyl-coenzyme A thioesterase, Acetyl-CoA acetyltransferase, Pyruvate carboxylase, Peroxiredoxin-6 et Fatty acid-binding protein). Des adipokines comme l'adiponectine et la leptine qui rendent compte de la fonction endocrine des adipocytes ont été identifiées et leurs abondances augmentent à partir de 180 jpc. Toutes ces protéines ont une abondance qui augmente entre 110 et 260 jpc et surtout à partir de 180 jpc. L'âge de 180 jpc semble être une phase de transition importante entre la prolifération et la différenciation.

En outre, cette analyse du protéome a révélé de manière surprenante des protéines marqueurs d'adipocytes blancs (comme LEP, UCP2 ATP5B, ALDH2, ALDH9A1 et SELENBP1 ; (Cinti *et al.*, 1997; Digby *et al.*, 2000)) alors que des marqueurs d'adipocytes bruns (comme DIO2 et UCP1 ; (Cinti *et al.*, 1997; Giralt *et al.*, 1989)) étaient attendus. L'hétérogénéité cellulaire des TA fœtaux a ensuite été confirmée par l'identification histologique d'adipocytes blancs et bruns dès 180 jpc à la fois dans le TA périrénal (Taga *et al.*, 2011), mais aussi dans le TA intermusculaire (Taga *et al.*, 2012).

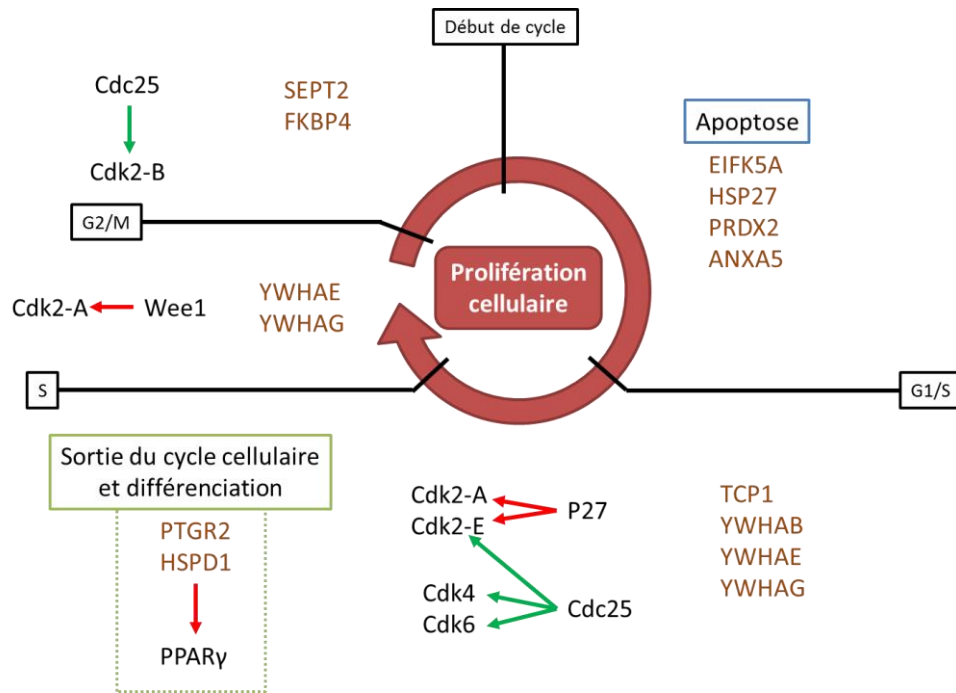


Figure 11 : Protéines potentiellement impliquées dans la prolifération des progéniteurs adipeux (adapté de la thèse d’Hager Taga, (Taga, 2011)). Les flèches vertes représentent les activations et les flèches rouges les inhibitions du cycle cellulaire. Les protéines en marron ont été identifiées par Taga et al. (Taga et al., 2012) et ont un rôle potentiel dans la prolifération des précurseurs adipocytaires par le contrôle de la régulation du cycle cellulaire, de l’apoptose ou du retardement de la différenciation induit par PPAR γ .

Les facteurs de variations de l’adipogénèse chez le bovin restent peu connus. Le facteur de variation le plus étudié est la race. En fin de vie foetale, il n’y a pas de différences entre races pour le poids de TA périrénal lorsque les races Frisonne et Charolaise (Robelin, 1986), Angus et Brahman (Landis et al., 2002) ou Holstein, Blonde d’Aquitaine et Charolaise (Taga et al., 2011) sont comparées. Par contre, le TA périrénal de fœtus Holstein possède moins d’adipocytes mais ils sont plus gros du fait de plus fortes activités lipogéniques que ceux des TA de fœtus Blond d’Aquitaine et Charolais (Taga et al., 2011). Enfin, lorsque les protéomes des TA de fœtus Blond d’Aquitaine et Charolais ont été comparés, seule l’abondance de 5 protéines variait (Taga et al., 2012).

3 – Interactions entre les tissus musculaires et adipeux

Les connaissances sur la croissance des tissus suggèrent une priorisation ou une compétition pour la croissance des muscles comparativement à celle des TA. En effet, la croissance du TA augmente quand celle du muscle diminue. La comparaison de génotypes extrêmes (bovins culards *versus* vaches laitières par exemple) indique que les races bovines diffèrent par leur aptitude de développement des TA et des muscles. Les races précoces (Angus, Noir Japonais, Holstein dont la croissance musculaire s’atténue rapidement avec l’âge) ont une proportion de TA corporel plus élevée (Gotoh et al., 2009) que les races tardives (Blanc Bleu Belge, Blonde d’Aquitaine, Limousin, au potentiel de croissance musculaire plus élevé). Comparativement aux races tardives, les races précoces ont en général un nombre d’adipocytes plus élevé, des adipocytes plus gros, des activités

enzymatiques pour le dépôt de lipides plus fortes et des fibres musculaires, en particulier les fibres rapides glycolytiques, plus petites. Des bovins Blanc Bleu Belge, dont l'hypermuscularité résulte d'une mutation du gène de la myostatine, ont un NTF jusqu'à deux fois plus élevé dans certains muscles et moins de dépôts adipeux dans les muscles et le corps entier que des génotypes non mutés (Bonnet *et al.*, 2010). Ainsi, des co-régulations affectant soit le nombre soit le volume des fibres musculaires comparativement aux adipocytes, se succéderaient et/ou seraient concomitantes selon le stade de développement des tissus. Les co-régulations qui affectent le nombre de cellules musculaires et adipeuses surviendraient essentiellement durant la vie fœtale et avant l'âge auquel le NTF est fixé, soit environ 180 jpc chez le bovin. Les co-régulations qui affectent le volume des cellules adipeuses et musculaires surviendraient tout au long de la vie de l'animal.

3.1 – Des interactions pour la régulation du nombre de cellules musculaires et adipeuses

Durant les phases précoces de la croissance, 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 (NTF). Ces variations apparaissent 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 et de l'espèce (en liaison avec la présence ou pas de TA brun). Une mobilisation différentes de ces précurseurs selon la race ou l'environnement intra-utérin pourrait modifier l'équilibre entre cellules musculaires et adipeuses et faire intervenir ou non des mécanismes d'interactions entre tissu.

Les mécanismes d'interactions entre les tissus adipeux et musculaires durant les phases précoces de développement ou durant la croissance des tissus par hyperplasie, restent largement à décrire. Ils sont cependant fortement soutenus par des résultats expérimentaux chez les rongeurs. Par exemple, chez la souris l'ablation génique de cellules musculaires favorise la différenciation des adipocytes musculaires. Inversement, l'ablation génique de cellules adipeuses altère la régénération du muscle blessé (Liu *et al.*, 2012). Certains mécanismes d'interactions orientent probablement les progéniteurs adipo-myogéniques vers le lignage adipeux ou musculaires (Figure 12 ; (Bianco *et al.*, 2011; Bonnet *et al.*, 2015; Levi et Longaker, 2011)). En effet une cellule mésenchymateuse issue du dermomyotome serait à l'origine du lignage musculaire. Elle donnerait lieu à un progéniteur qui exprimerait Myf5+ et qui se différencie non seulement en fibres musculaires mais aussi en adipocytes (Bonnet *et al.*, 2015). En effet, les cellules Myf5+ seraient à l'origine des adipocytes blancs et bruns dont la proportion varie en fonction de la localisation anatomique des TA chez la souris (Sanchez-Gurmaches et Guertin, 2014). Les cellules Myf5+ seraient à l'origine de la majorité des adipocytes du TA brun interscapulaire, des TA sous-cutanés bruns de la partie haute du corps et du TA blanc rétro-péritonéal (Figure 13). Par contre, ils contribueraient très peu à la formation des TA intermusculaire, périgonadique et inguinal chez la souris (Sanchez-Gurmaches et Guertin, 2014). La détermination du progéniteur Myf5+ vers le lignage musculaire est sous le contrôle des régulateurs transcriptionnels Myf5, MyoD, MRF4 (Bonnet *et al.*, 2010; Park *et al.*, 2014). La détermination du

progéniteur Myf5+ vers le lignage adipeux dépendrait de BMP7 et PRDM16 (Tang et Lane, 2012; Timmons et al., 2007; Tseng et al., 2008). La différenciation des progéniteurs Myf5+ en préadipocytes bruns dépendrait de PPAR γ et α , CCAAT, C/EBP- β et PGC-1 α et β (Park et al., 2014). Le rôle central de PRDM16 dans l'orientation des progéniteurs adipo-myogéniques vers le lignage adipeux brun a été révélé *in vitro*. Son inhibition dans des progéniteurs adipeux bruns induit leur différenciation en myoblastes du fait de la surabondance de facteurs myogéniques (Seale et al., 2008). La surexpression de PRDM16 par des myoblastes oriente leur phénotype vers celui des adipocytes bruns (Seale et al., 2008). D'autres protéines telles que les membres de la superfamille des transforming growth factor- β (TGF- β) seraient aussi impliquées dans l'orientation de précurseurs adipo-myogéniques vers le lignage adipeux ou musculaire (Kawaguchi et al., 2010). En effet, la myostatine (Amthor et al., 2002) et l'activine A (He et al., 2005) inhibent le développement musculaire durant l'embryogenèse chez le poulet en inhibant l'expression de Pax3 et de Myf5. Or, ces protéines favorisent la différenciation vers le lignage adipeux de progéniteurs de type C3H10T1/2 (Artaza et al., 2005) et adipo-myogéniques humains (Zaragosi et al., 2010). Enfin, la voie des Wnt pourrait participer aux co-régulations de la myogenèse et de l'adipogenèse. En effet, l'activation de la voie Wnt inhibe la différenciation adipocytaire mais favorise la différenciation myogénique (Seale et al., 2009). En particulier, la glycoprotéine Wnt10b inhibe l'adipogenèse *in vitro* dans les adipocytes bruns et blancs chez les monogastriques en bloquant l'induction de facteurs de transcription de l'adipogenèse (PPAR γ et PGC1 α ; (Seale et al., 2009)). Cet effet inhibiteur a aussi été rapporté *in vitro* chez le bovin (Tan et al., 2006). L'action inhibitrice de la myostatine sur l'adipogenèse des adipocytes blancs résulte d'ailleurs d'une activation de la voie de signalisation Wnt (Guo et al., 2008).

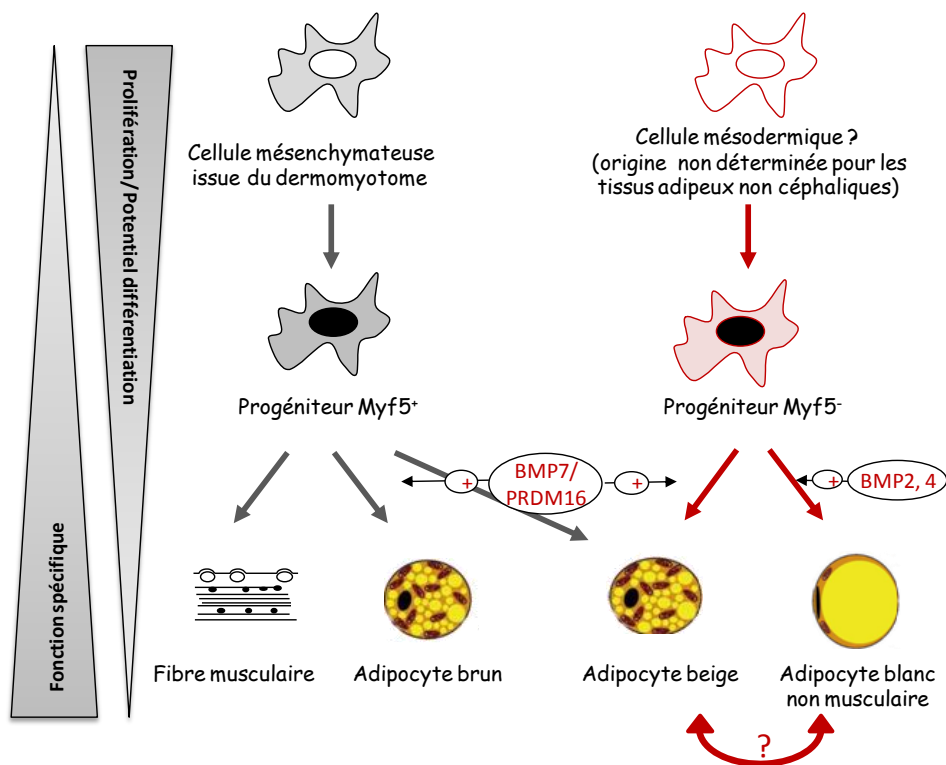


Figure 12 : Cellules à l'origine des lignages des précurseurs musculaires et adipeux suivant les activateurs de leur détermination cellulaire (Bonnet et al., 2015).

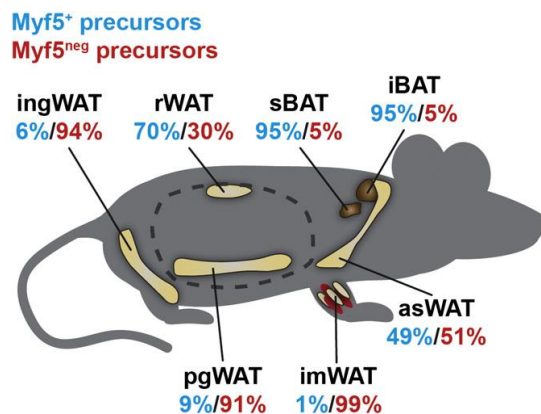


Figure 13: Les deux types de progéniteurs Myf5, dans chaque dépôt de TA, varient en fonction de la localisation anatomique (Sanchez-Gurmaches et Guertin, 2014). On y observe les TA blancs (WAT) et brun (BAT) de différentes localisations anatomiques : le TA inguinal (aine ; ingWAT), rétropéritonéal (rénal ; rWAT), sous-scapulaire (sous l'épaule ; sBAT), interscapulaire (entre les épaules ; iBAT), périgonadal (entourant l'utérus et les ovaires chez les femelles, l'épididyme et les testicules chez les mâles ; pgWAT), mésentère (surface des intestin grêle ; imWAT) et sous-cutané (asWAT).

Enfin, des interactions entre les tissus adipeux et musculaires pourraient impliquer des protéines sécrétées par ces tissus, les myokines (Pedersen, 2011) pour le muscle et les adipokines pour le TA (Trayhurn *et al.*, 2006). Les adipo-myokines sont synthétisées par les deux tissus (Raschke et Eckel, 2013; Trayhurn *et al.*, 2011). Ces protéines sont capables d'agir au sein d'un tissu (paracrine et/ou autocrine) ou *via* la circulation sanguine (endocrine). Comme le muscle se développe plus précocement que le TA, les myokines (Pedersen, 2011; Pedersen *et al.*, 2007) sont susceptibles de moduler l'adipogenèse. Parmi les myokines, l'interleukine 6 (IL-6) (Steensberg *et al.*, 2000), est connue pour inhiber la différenciation adipocytaire en agissant sur la voie de signalisation Wnt (Gustafson et Smith, 2006).

3.2 – Des interactions pour la régulation du volume des cellules musculaires et adipeuses

L'hypertrophie des cellules musculaires et adipeuses est l'événement cellulaire majeur à l'origine des croissances fœtales et surtout post-natales des TA et des muscles. Durant ces phases de croissance une interaction négative entre TA et muscles 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 ou le glucose. Au niveau du corps entier, un exemple 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 *et al.*, 2002). L'orientation des nutriments vers l'accumulation de protéines, lipides ou glycogène dans les fibres musculaires ou vers le dépôt de lipides dans les adipocytes est sous le contrôle de nombreux nutriments, hormones et facteurs de croissance (Pethick et Dunshea, 1996; Sillence, 2004) en interactions avec les besoins en nutriments

de chaque tissu/cellule et avec la disponibilité en ces nutriments (liée au statut nutritionnel et physiologique de l'animal). En outre, les muscles et les TA synthétisent et sécrètent de nombreuses protéines : les myokines comme la myostatine ([Artaza et al., 2005](#); [Hirai et al., 2007](#)) et les adipokines comme la leptine ([Poulos et al., 2010](#)) capables d'agir par voie endocrine ou paracrine et de contribuer au dialogue entre adipocytes et fibres musculaires. Outre ces concepts généraux largement partagés, à ce jour les interactions restent peu décrites. Nous illustrerons deux types d'interactions entre les TA et les muscles : des interactions faisant intervenir des métabolites et des protéines qui régulent le partage des nutriments, et des interactions impliquant la sécrétion de régulateurs de la croissance des tissus. Les deux tissus sont connus pour synthétiser et sécréter un grand nombre de peptides, appelés adipokines pour le TA ([Trayhurn et al., 2006](#)) et myokines pour le muscle ([Pedersen, 2011](#)). Elles agissent soit au sein de la cellule (actions paracrine et/ou autocrine), soit à distance du fait de leur sécrétion dans les liquides physiologiques (endocrine). Des protéines sont sécrétées à la fois par le TA et le muscle, parfois pour agir d'un tissu sur l'autre et sont appelés des adipo-myokines ([Raschke et Eckel, 2013](#); [Trayhurn et al., 2011](#)).

Parmi les interactions qui font intervenir des nutriments, les acides gras non estérifiés (AGNE) sécrétés par le TA en réponse à une mobilisation/lipolyse sont capables de réduire l'utilisation du glucose. Deux voies de signalisation sur le muscle ont été décrites. Elles conduisent à une diminution de l'abondance du transport de glucose (GLUT4) du fait l'augmentation intracellulaire du citrate ([Randle, 1956](#)) et du fait de l'activation de serine/threonine kinases telles que la protéine kinase C ([Morino et al., 2006](#)).

Pour les myokines qui participent aux interactions entre les muscles et TA, l'interleukine-6 sécrétée par le muscle en contraction, agit sur le TA et notamment le TA intramusculaire en activant la lipolyse ([Pedersen et Fischer, 2007](#)). L'interleukine-15 joue un rôle dans la réduction de la masse du TA ([Pedersen et al., 2007](#)). L'irisine est une myokine sécrétée par le muscle en réponse à l'exercice et qui régule positivement la différenciation des adipocytes blancs ([Raschke et al., 2013](#)). En outre, l'irisine est présente dans le muscle bovin et son abondance n'est pas liée à la composition corporelle ([Komolka et al., 2014](#)). Par contre contrairement à la souris, la protéine n'a pas été identifiée dans le sérum de bovins ([Komolka et al., 2014](#)). Enfin, la myonectine augmente le prélèvement des acides gras par le TA blancs chez la souris, réduisant ainsi les taux d'acides gras circulants ([Seldin et al., 2012](#)).

Via les adipokines, le TA est maintenant reconnu comme un tissu hautement sécréteur qui participe ainsi aux dialogues avec de nombreux autres tissus et d'autres organes pour agir sur le métabolisme énergétique, la reproduction, la mitose, l'angiogenèse ou encore l'adipogenèse par exemple ([Kawwass et al., 2015](#); [Woronuk et al., 2012](#)). L'exemple le mieux décrit chez le ruminant concerne la leptine ([Chilliard et al., 2001](#); [Chilliard et al., 2005](#)). L'un des rôles essentiels de la leptine est d'assurer l'équilibre des réserves lipidiques en régulant la prise alimentaire et la dépense énergétique, notamment l'oxydation des acides gras. Chez les bovins, une forte adiposité est concomitante avec une augmentation de la teneur plasmatique en leptine accompagnée, au niveau du muscle, d'une augmentation de l'oxydation des acides gras et d'une diminution de la lipogenèse ([Bonnet et al., 2010](#)). Ainsi chez le bovin, comme chez les monogastriques, la leptine favoriserait l'utilisation des acides gras dans le muscle plutôt que leur stockage dans le tissu adipeux. D'autres adipokines telles que l'adiponectine auraient des effets similaires d'augmentation de l'oxydation des

acides gras dans la cellule musculaire, stimulant ainsi la sensibilité à l'insuline chez l'homme ([Dyck, 2009](#))

Ainsi pour comprendre les interactions entre le TA et le muscle, il est maintenant nécessaire d'avoir une bonne connaissance des activités de sécrétions des TA et des muscles. De telles connaissances sont disponibles chez les monogastriques (voir paragraphe 4). La fonction de sécrétion du TA a commencé à être étudiée chez le bovin avec les études sur les facteurs de variation de la synthèse de leptine en réponse à l'alimentation ([Bocquier et al., 1999](#); [Chilliard et al., 2000](#)), l'adiposité ([Delavaud et al., 2000](#); [Delavaud et al., 2002](#); [Sadri et al., 2011](#); [Yonekura et al., 2013](#)) ou l'âge foetal ([Taga et al., 2012](#)). La fonction de sécrétion continue à être étudiée en ciblant une adipokine comme l'adiopnectine ([Lemor et al., 2010](#); [Taga et al., 2012](#)), la visfatine et la résistine ([Kratchmarova et al., 2002](#); [Lemor et al., 2010](#)). Malheureusement, contrairement aux monogastriques, à ce jour aucune étude de génomique n'a été conduite pour décrire plus largement la capacité de sécrétion des TA et des muscles chez les ruminants.

4 - Intégration de données publiques pour une meilleure compréhension de l'adipogenèse, de la myogenèse et des interactions entre le TA et le muscle

La synthèse des données disponibles révèle une mauvaise connaissance des mécanismes précoces de la myogenèse, de l'adipogenèse et en particulier de la détermination et de la prolifération des progéniteurs musculaires et adipeux. En outre, bien que suggérées depuis longtemps, les interactions entre les tissus musculaires et adipeux restent largement à décrire. Or, de nombreuses études de génomique publiées et stockées dans des bases de données publiques, telles que Gene Expression Omnibus (GEO ; [Barrett et al., 2013](#)) et ArrayExpress ([Kolesnikov et al., 2015](#)), rapportent des transcrits ou des protéines dont les abondances varient en fonction des étapes de la myogenèse ou de l'adipogenèse. L'hypothèse que nous faisons est que la synthèse des données disponibles serait une stratégie pour une meilleure compréhension des étapes précoces de l'adipogenèse, de la myogenèse et des interactions entre le TA et le muscle. C'est dans cet objectif que la base de données Fat&MuscleDB a été créée ([Tournayre et al., 2014](#); [Tournayre et al., 2015](#)). Il s'agit d'une base de données qui regroupe les résultats de transcriptomique et de protéomique d'expérimentations *in vivo* et *in vitro* chez les ruminants, mais aussi chez les espèces modèles ou des lignées cellulaires lorsque les données n'étaient pas disponibles chez le ruminant. En novembre 2015, Fat&MuscleDB regroupait les données d'environ 180 publications et 100 jeux de données GEO sélectionnés pour leur pertinence quant à la compréhension de la myogenèse, de l'adipogenèse et des interactions entre ces 2 processus.

Parmi ces données certaines pourraient permettre une meilleure compréhension des mécanismes précoces de l'adipogenèse ou de la myogenèse, notamment ceux de la détermination ou la prolifération cellulaire ([Tableau 3](#)). Il s'agit des données d'abondances des transcrits et des protéines qui varient selon l'état de prolifération et de différenciation de cellules souches pluripotentes telles que la lignée C3H10T1/2 (18600 identifiants (ID) uniques), ou issues de tissus musculaires (23600 ID uniques) et adipeux (19900 ID uniques). Une synthèse des données disponibles sur les lignées modèles de l'adipogenèse, comme les cellules 3T3-L1 (24600 ID uniques)

et les lignées de la myogenèse, telles que les cellules C2C12 (27100 ID uniques), pourrait révéler des gènes clés dans la prolifération de ces types cellulaires. L'essentiel de ces données provient d'expérimentation *in vitro*, à partir de lignées ou de cultures primaires de cellules de monogastriques. Or, ce type de données *in vitro* n'existent pas chez les ruminants. Aussi les données *in vitro* produites chez les monogastriques pourraient être synthétisées pour identifier des gènes ou protéines retrouvés dans différentes conditions et ces données pourraient être confrontées aux données *in vivo* obtenues durant les phases précoces de développement, c'est-à-dire pendant la vie foetale pour les grands mammifères. Cependant, ce type de données est extrêmement rare ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Lehnert et al., 2007](#); [Romao et al., 2014](#); [Sudre et al., 2003](#); [Taga et al., 2012](#)).

Parmi les données de Fat&MuscleDB ([Tableau 3](#)), certaines pourraient favoriser la découverte de gènes ou des protéines clés dans l'hypertrophie des tissus musculaires et adipeux, car la variabilité de leur expression ou de leur abondance sera systématiquement liée à une variabilité de la masse des tissus. L'intégration des données chez le bovin suggèrent qu'environ 15330 et 2300 ID uniques sont associés respectivement, à une hypertrophie/atrophie des muscles et TA. Chez les espèces modèles telles que l'homme, le rat et la souris, la même intégration de données suggère qu'environ 16080 et 35590 ID uniques sont associés à une hypertrophie/atrophie respectivement des muscles et TA. Cette synthèse, bien que non exhaustive, met en évidence le manque de données de génomique dans le TA bovin qui représente seulement 6% des données TA des autres espèces.

Un rapide aperçu des données issues d'étude du sécrétome du muscle et du TA ([Tableau 3](#)) montre qu'à ce jour aucune n'est disponible chez le bovin. Pour les espèces modèles, on référence 1940 et 2070 ID uniques pour les données liées respectivement aux sécrétomes du muscle et du TA. Ces données pourraient apporter des nouvelles pistes d'études pour la régulation de la croissance des tissus et des interactions entre tissus qui impliquent des myokines et des adipokines.

Ce recensement rapide et non exhaustif illustre que de nombreuses données sont déjà disponibles et pourraient permettre une meilleure connaissance de la myogenèse, de l'adipogenèse et des interactions entre ces deux processus. Utiliser uniquement les données de l'espèce bovine reviendrait à se passer des deux tiers des connaissances disponibles dans d'autres espèces, souvent mieux renseignées puisque très peu de données sont disponibles concernant l'adipogenèse et la myogenèse chez le bovin. Il sera nécessaire de confronter les connaissances obtenues chez le bovin à celles disponibles dans d'autres espèces afin de comprendre les mécanismes de mise en place des tissus musculaires et adipeux. Cependant, **l'analyse de cette masse de données requiert des outils informatiques pour identifier les fonctions biologiques des gènes et protéines impliqués et aussi pour étudier ou prédire le sécrétome des tissus.**

<i>In vitro</i>		
Type cellulaire	Toutes espèces	Bovin
Lignée de myoblaste C2C12	27100	0
Lignée de préadipocytes 3T3L1	24600	0
Lignée pluripotente adipo-myogénique C3H10T1/2	18600	0
Cellule souche adulte myogénique	23600	30
Cellule souche adulte adipogénique	19900	20
Culture primaire de progéniteurs musculaires	37400	0
Culture primaire de progéniteurs adipocytaires	45300	0
Nombre d'identifiants pour les données <i>in vitro</i> liées au muscle	80600	30
Nombre d'identifiants pour les données <i>in vitro</i> liées au TA	73900	20
<i>In vivo</i>		
Phénotype de croissance	Toutes espèces	Bovin
Hypertrophie musculaire d'origine génétique	5350	15320
Hypertrophie musculaire induite par alimentation	30	4360
Atrophie musculaire induite par alimentation	14020	4370
Hypertrophie adipeuse d'origine génétique	50	1630
Hypertrophie adipeuse induite par alimentation	29560	770
Atrophie adipeuse induite par alimentation	31560	50
Nombre d'identifiants pour les données <i>in vivo</i> liées au muscle	16080	15330
Nombre d'identifiants pour les données <i>in vivo</i> liées au TA	35590	2300
Sécrétome		
Sécrétome	Toutes espèces	Bovin
Données liées au sécrétome du muscle	1940	0
Données liées au sécrétome du TA	2070	0
Nombre total d'identifiants	Toutes espèces	Bovin
Nombre total d'identifiants pour les données liées au muscle	81700	15330
Nombre total d'identifiants pour les données liées au TA	81970	2300

Tableau 3 : Recensement des protéines et transcrits identifiés pour 107 GSE et 183 publications scientifiques en rapport avec l'adipogenèse ou la myogenèse (novembre 2015). Ce recensement a été réalisé grâce à l'outil Fat&MuscleDB ([Tournayre et al., 2015](#)) pour des expérimentations *in vitro* et *in vivo* chez le bovin, l'homme, le rat, la souris, le porc et le mouton.

5 – La fouille de données de génomique par bioinformatique

On pourrait qualifier d'étude bioinformatique, n'importe quelle étude qui tend à numériser une expérimentation biologique et/ou les résultats qui en découlent. Ainsi, la bioinformatique est l'étude de faits biologiques grâce à des représentations informatiques ([Luscombe et al., 2001](#)). L'objectif d'une démarche bioinformatique est d'apporter de nouvelles connaissances ou hypothèses biologiques. La bioinformatique possède de nombreux domaines d'application, comme l'analyse de données de génomique (transcriptomique et protéomique), la représentation en 3D de la structure des molécules, la fouille de données (le datamining dans la littérature ou dans des bases de données) ou encore la recherche de ligands/récepteurs entre molécules pour l'immunologie ou la pharmacologie ([Berglund et al., 2009](#); [Chen et Chen, 2008](#); [Kumar et Dudley, 2007](#); [Schmidt et al., 2014](#); [Sleator et Walsh, 2010](#)). Une revue complète de l'état actuel de la recherche en bioinformatique est impossible, car des centaines de nouveaux outils et bases de données sont créés chaque année ([Cannata et al., 2005](#); [Stockinger et al., 2014](#); [Wanichthanarak et al., 2015](#)). Aussi, ce chapitre portera sur les stratégies et outils bioinformatiques les plus communément utilisés pour la fouille de données de génomique.

La génomique a pour but d'identifier et de quantifier l'expression d'un maximum de gènes en mesurant l'abondance de transcrits (transcriptomique) ou de protéines (protéomique) à un moment donné et dans des conditions données au sein d'un tissu, d'une cellule ou d'un compartiment cellulaire. L'objectif des études de génomique est de mettre en évidence, pour une condition expérimentale étudiée, les gènes et les protéines dont les expressions et les abondances sont modifiées et de comprendre quelles fonctions biologiques ils impactent. Comme cette fouille de données concerne souvent plusieurs centaines d'identifiants, une assistance informatique est nécessaire. Ce paragraphe présente les différents types de données de génomique et les stratégies communément utilisées pour les analyser.

5.1 – Disponibilité des données de génomique

Les données de génomique à analyser ont deux origines essentielles : des expérimentations de laboratoires ou des bases de données. Les expérimentations sont des analyses du transcriptome ou des analyses du protéome qui ont largement été décrites ([Haoudi et Bensmail, 2006](#); [Schmidt et al., 2014](#); [Vickers et al., 2015](#)). Les résultats de ces expérimentations sont souvent disponibles en ligne, grâce à des bases de données dédiées. Les plus connues et qui concernent les protéines ou la protéomique sont la base « Protein » du NCBI ([Coordinators, 2015](#)), ProteomeDB ([Wilhelm et al., 2014](#)) ou encore Peptide Atlas ([Deutsch, 2010](#)). La base de données européenne des protéines, UniProt ([UniProt, 2015](#)), a récemment mis en ligne une nouvelle section proposant les protéomes complets de centaines d'espèces (eucaryotes ou procaryotes). Les données de transcriptomique sont disponibles sur des bases de données comme Gene Expression Omnibus (GEO ; [Barrett et al., 2013](#)). Ainsi, avant de commencer des études de génomique onéreuses et longues il serait préférable de vérifier en ligne si des données existantes pourraient répondre à la question biologique posée. Les données de génomique ont plusieurs formats : il peut par exemple s'agir de séquences au format FASTA, d'identifiants de protéines ou de gènes, ou encore de listes d'abondances de transcrits ou de protéines.

5.2 – Stratégie d’analyse pour une interprétation fonctionnelle de listes de gènes ou de protéines

Nous n’évoquerons pas les stratégies informatiques et statistiques utilisées pour déclarer des abondances différentielles ou pour identifier des groupes de gènes ou de protéines similaires par les variations de leurs abondances selon les conditions étudiées. Nous présentons ici les stratégies d’analyses pour une interprétation des fonctions biologiques liées à des listes d’identifiants géniques ou protéiques. Globalement, les stratégies d’analyses sont identiques pour les listes d’identifiants issus d’expérimentations de transcriptomique ou de protéomique. L’objectif de ces analyses est d’identifier les fonctions biologiques les plus représentatives associées à une liste d’identifiants, avec comme cadre d’analyse la question biologique au regard des connaissances disponibles. Le résultat de ces analyses est de proposer une hypothèse pouvant répondre à une question de recherche, l’hypothèse pouvant ensuite être validée expérimentalement. La stratégie d’analyse la plus simple est de compiler la bibliographie disponible pour chaque identifiant. Cependant, avec l’augmentation de la sensibilité des méthodes de détections, notamment en protéomique, les listes de gènes et de protéines sont maintenant constituées de centaines d’identifiants. Aussi, l’analyse des identifiants un à un est un travail fastidieux et peu productif. Ainsi, la bioinformatique est d’une aide précieuse pour les biologistes, car elle propose un large panel d’outils pour l’analyse de ces données : allant de l’analyse de séquences protéiques à la recherche d’interactions entre des protéines. Ces outils sont des bases de données publiques, qui recensent les nouvelles connaissances biologiques (molécules, fonctions biologiques, localisations cellulaires, les interacteurs, les co-régulations, les méthodes expérimentales...), mais aussi des outils dédiés à l’analyse de données (workflows, web services ou standalone pour le traitement de données et la prédiction d’information) ([Brazas et al., 2011](#); [Lisacek et al., 2006](#)).

Lisacek et al. ([Lisacek et al., 2006](#)), Laukens et al. ([Laukens et al., 2015](#)) et Schmidt et al. ([Schmidt et al., 2014](#)) ont proposé des stratégies d’analyse pour des listes de protéines ou de gènes. De ces stratégies, un processus commun en plusieurs étapes émerge ([Figure 14](#)). La première étape est d’interroger les bases de données pour annoter les identifiants de la liste grâce à des informations pour chaque protéine ou gène telles que le nom, la fonction biologique, le tissu d’expression, les isoformes, des références bibliographiques majeures... Cette première partie a pour objectif de synthétiser les connaissances disponibles pour chaque identifiant d’une liste sans avoir à compiler la littérature existante. Il existe de nombreuses bases de données qui rassemblent les connaissances disponibles pour une protéine ou un gène, les plus communément utilisées et mises à jour sont UniProt Knowledge Base ([UniProt, 2015](#)), NCBI (National Center for Biotechnology Information ; ([Coordinators, 2015](#))) et Ensembl ([Cunningham et al., 2015](#)). La seconde étape a pour objectif de réduire la redondance des termes associés aux processus biologiques. Pour cela, un identifiant est lié à des termes relatifs aux fonctions biologiques et moléculaires ou à la localisation cellulaire, ceci afin de structurer la description des gènes et des protéines dans le cadre d’une ontologie commune à toutes les espèces (appelée Gene Ontology, GO). Ensuite, des études d’enrichissement identifient les termes GO les plus fréquemment trouvés pour annoter la liste des identifiants. Il existe de nombreuses bases de données qui annotent des gènes et des protéines avec des termes GO. La plus communément utilisée est la base du Gene Ontology Consortium ([Gene Ontology, 2015](#)). La troisième étape est de lier les identifiants d’un jeu de données entre eux, on parle de réseaux moléculaires, afin de mettre en évidence des réseaux d’identifiants contribuant à

une fonction biologique (activation de fonction par exemple), d'identifier des interactions entre les identifiants d'un jeu de données...La visualisation des liaisons ou interactions entre identifiants, est souvent réalisée sous la forme de réseaux. Des observations topologiques simples des réseaux, comme le nombre de liaisons entre deux identifiants, mettent en évidence les identifiants ayant une fonction centrale dans la liste.

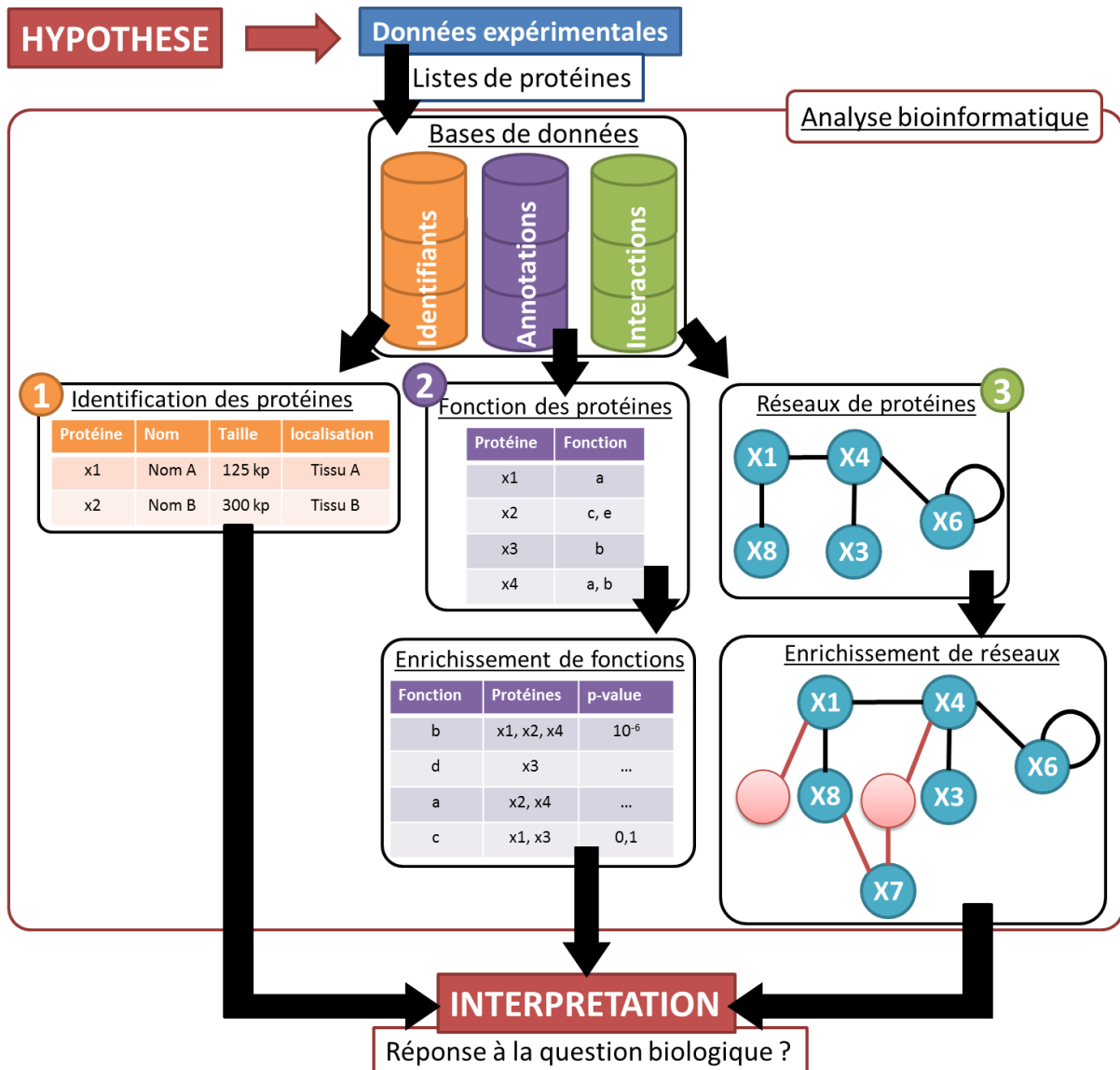


Figure 14 : Stratégie générale d'analyse de listes de protéines ou gènes issues d'études de protéomique (adapté de (Laukens et al., 2015)).

Outre cette stratégie générale appliquée à des listes de gènes ou de protéines, il existe une multitude d'autres algorithmes pour une analyse assistée par bioinformatique (Henry et al., 2014). Dans le cas de la recherche de protéines susceptibles d'intervenir dans le dialogue entre tissus, tel que celui que nous avons évoqué pour les TA et le muscle (cf. paragraphe 3), une stratégie est d'identifier les protéines sécrétées (Forrester et al., 2014; Prassas et al., 2012). Certaines bases de

données comme UniProt renseignent les protéines connues pour être sécrétées. Il existe des bases spécifiques aux protéines sécrétées comme Metazoa, une base de données de sécrétome et de localisation cellulaire ([Meinken et al., 2015](#)). Il est aussi possible de prédire les protéines potentiellement sécrétées, par la recherche de motif sur leur séquence protéique ([Meinken et Min, 2012](#)). À titre d'exemple, la suite d'outils du CBS (Center of Biological Sequence Analysis) permet de prédire les protéines potentiellement sécrétées dans de nombreuses espèces procaryotes, animales ou végétales (<http://www.cbs.dtu.dk/services/>).

Une fois toutes les données analysées, l'interprétation doit permettre d'apporter des éléments de réponse à l'hypothèse de recherche ou de formuler une nouvelle hypothèse. Pour cette stratégie d'analyse souvent décrite, de nombreux outils et bases de données dédiés à l'analyse de données de génomique existent. Le projet OmicTools tente de rassembler le maximum d'outils et bases de données sur une même plateforme. En novembre 2015, il référençait déjà plus de 11 000 ressources ([Henry et al., 2014](#)). OmicTools propose pour chaque ressource une fiche descriptive, une publication, les avis et questions des utilisateurs. Nous avons réalisé une revue non exhaustive des outils et bases de données disponibles et couramment utilisés pour étudier l'analyse des données de génomique selon la stratégie décrite dans la [Figure 14](#). Les ressources disponibles et dédiées à l'analyse de données de ruminants sont aussi illustrées.

6 – Revue des outils bioinformatique pour l'analyse de données de génomique

6.1 – Les bases de données référençant les protéines et les gènes

Une base de données est une plateforme de stockage dans le cas présent de l'information biologique connue, par exemple : la transcriptomique, la protéomique, la phylogénie des espèces, des méthodes expérimentales ou des publications de journaux scientifiques. Le rapport annuel du journal Nucleic Acids Research « Molecular Biology Database Collection » dénombre autour de 1550 bases de données biologiques référencées ([Galperin et al., 2015](#)). Certaines bases sont accessibles gratuitement et donnent un libre accès à l'utilisation de leurs données par exemple la base GeneBank ([Benson et al., 2015](#)). D'autres bases sont payantes et proposent un accès limité à la source de leurs données (« Ingenuity Pathway Analysis » par exemple ; <http://www.ingenuity.com/products/ipa>). Une revue récente propose une collection de 114 bases de données dédiées aux différents domaines de recherche chez l'homme ([Zou et al., 2015](#)).

Les informations stockées dans les bases de données biologiques ont plusieurs formes. Ce peut être des informations textuelles (ce sont des faits ou des descriptions) : identifiants biologiques, séquences (ADN, ARN et protéiques), les liens entre des molécules (directs, indirects, cascade de réactions...), et les articles scientifiques par exemple. Des bases proposent une hiérarchisation des informations : ontologie des gènes et phylogénie des espèces. Certaines bases proposent des informations sous forme de schémas spatiaux : structures 3D de molécules (coordonnées spatiales des atomes) ou localisations chromosomiques des gènes. Des bases référençant les expressions des gènes avec les données de puces à ADN et ARN. D'autres bases sont construites comme des banques de données : il s'agit de fichiers structurés composés d'informations collectées autour d'une thématique particulière, par exemple toutes les séquences ADN du bovin au format fasta. Il faut aussi

préciser qu'une base de données biologiques possède ses propres identifiants pour chaque gène et protéine, ce qui oblige souvent les utilisateurs à convertir leurs données au format reconnu par la base ou l'outil à interroger ([Flicek et al., 2014](#)). Par exemple, les différents identifiants associés à l'adiponectine : Q15848 (UniprotKB accession), ADIPO_HUMAN (UniprotKB Symbol), 9370 (Gene ID de NCBI), ADIPOQ (Gene Symbol) ou encore ENSP00000320709 et ENSG00000181092 (respectivement identifiant protéique et génétique de la base Ensembl).

Les données répertoriées dans ces bases sont soumises à des contrôles d'authenticité, on parle de « curation » étymologiquement du latin *curare* : prendre soin et de l'anglais « content curation » ou « data curation ». Ces contrôles sont automatiques ou manuels et renseignent un utilisateur sur la qualité de l'information stockée. Ainsi, les bases de données proposent souvent une notation de l'authenticité de leurs données sous les formes suivantes :

- Reviewed (revue): information revue et corrigée par un membre du comité de révision de la base de données (validation manuelle, on parle de curation). Il a pu y ajouter des informations provenant de publications scientifiques et de méthodes d'obtention de l'information.
- Validated (validée): l'information a été acceptée par la base de données (validation électronique), mais elle peut encore nécessiter une révision manuelle humaine pour une validation de meilleure qualité.
- Provisional (provisoire) : l'information vient d'être ajoutée, elle n'a pas encore été validée (électroniquement ou manuellement). L'information provient de nouvelles expérimentations en laboratoire.
- Predicted (prédite) : l'information entrée provient d'outils de prédiction et n'a pas été vérifiée expérimentalement. Elle peut faire l'objet d'une validation et d'une révision, mais l'information restera considérée comme prédite.

Souvent les bases de données proposent en plus un autre système d'authentification des données basé sur des calculs statistiques (*p*-value ou indice de confiance par exemple) pour orienter l'utilisateur vers les données les plus fiables.

Les bases de données biologiques en génomique

Pour rechercher des informations générales relatives à un gène ou à une protéine, les bases de données biologiques les plus communément utilisées ([Fernandez-Suarez et Galperin, 2013](#); [Galperin et Fernandez-Suarez, 2012](#)) sont : UniProt Knowledge Base ([UniProt, 2015](#)), NCBI ([Coordinators, 2015](#)), ExPASy ([Artimo et al., 2012](#)), EBI ([Brooksbank et al., 2014](#)) et KEGG ([Kanehisa et al., 2014](#)) (Tableau 4). Elles rassemblent des données de protéomique, mais aussi de transcriptomique et proposent aussi des outils d'analyse de leurs données. Ces bases de données sont libres d'utilisation, sont régulièrement mises à jour et proposent des versions téléchargeables de toutes leurs données. Nous avons détaillé le contenu des bases de données NCBI et UniProtKB qui sont indispensables à toute étude de génomique et de transcriptomique. D'autres bases de données répertoriées dans le [Tableau 4](#) sont présentées plus succinctement.

Bases de données	Gratuité d'accessibilité	Description	Composition générale	Mise à jour	Références (DOI)
UniProt	Oui	La mission de la base de données UniProt est de fournir à la communauté scientifique une synthèse globale de haute qualité sur la protéine, ses fonctions biologiques et des ressources librement accessibles pour des analyses	Fiche d'information sur les protéines : identifiants, fonction, ontologie, séquence, localisation cellulaire, tissu d'expression et bibliographie	Mensuelle	10.1093/nar/gku469
NCBI	Oui	Le Centre National d'Information sur les Biotechnologies (NCBI) fait progresser la science et la santé en fournissant un accès à l'information biomédicale et génomique.	Fiche d'informations (protéines, gènes, ARN...), outils (alignements, puces biologiques...), bibliographie, phylogénie et homologie d'espèces	Journalière	10.1093/nar/gkt1146
EBI	Oui	EBI est une base de données libre qui fournit des informations tirées d'expérimentations des sciences de la vie, effectuée de la recherche fondamentale en bioinformatique et offre un programme de formation des utilisateurs	Fiche d'informations, ontologie, outils (séquençage, ontologie, interactions moléculaires...), bibliographie, cours et tutoriaux d'analyse bioinformatique	Journalière	10.1093/nar/gkt1206
KEGG	Non (sauf académique)	KEGG est une base de données qui répertorie les fonctions, voies métaboliques ou voies de différents systèmes biologiques, comme la cellule, l'organisme ou l'écosystème, à partir de grands jeux de données moléculaires générés par le séquençage à haut débit et la génomique	Fiche d'informations, outils (mapper, alignement et exploration de génome), bases de données de voie métaboliques	Mensuelle	10.1093/nar/gkt1076
Reactome	Oui	Reactome est une plateforme composée d'outils et d'une base de données revue par des experts (données expérimentales et prédites). Elle fournit des outils d'étude du génome et permet de visualiser, interpréter et analyser des voies métaboliques et des réseaux biologiques.	Fiche d'informations (protéines, gènes et petites molécules/drogues) et outils (visualisation de voies métaboliques, réseaux d'interaction de protéines, comparaison de réseaux entre espèces)	Trimestrielle	10.1093/nar/gkt1102
GeneCards	Non (sauf académique)	GeneCards est une base de données qui fournit des informations sur les gènes humains annotés et prédits. Elle intègre automatiquement les données de 125 sources, dont des données de génomique, de transcriptomique, de protéomique, de génétique et clinique.	Fiche d'informations (protéines, gènes, ARN...) et outils (annotations, voies métaboliques, analyse de liste de gènes, cartographie chromosomique)	Trimestrielle	10.1093/database/baq020

Tableau 4 : Référencement et description générale des bases de données biologiques utilisées communément (nom de la base de données, utilisation gratuite, description, données répertoriées, qualité des mises à jour et dernière publication).

Pour une synthèse des connaissances disponibles pour une protéine, la base de données la plus utilisée est UniProt Knowledge Base. The Universal Protein Resource (UniProt) est en effet la base de données européenne qui répertorie toutes les informations relatives aux protéines de milliers d'organismes. Pour sa distribution d'août 2015, UniProt rassemble 50 millions de séquences de protéines pour environ 560 000 espèces (eucaryotes et procaryotes). Actuellement, 60% des séquences entrées proviennent d'espèces procaryotes et 30% d'espèces eucaryotes. Seulement 550 000 de ces séquences ont été revues (manuellement ou automatiquement). En octobre 2014, UniProt a effectué une refonte complète de sa base de données et a amélioré son système de curation en ajoutant un système de notation en 5 points pour chacun des identifiants. Cette nouvelle notation concerne la qualité de l'expertise effectuée pour une protéine. Ainsi, chaque protéine répertoriée dans la base UniProt possède une fiche de renseignement composée : d'identifiants (de gènes et de protéines, et leurs synonymes sur d'autres bases de données), de sa fonction biologique dans l'espèce considérée (ainsi que toutes les autres fonctions dans lesquelles elle peut jouer un rôle), des annotations fonctionnelles, des séquences protéiques et d'autres informations (structure 3D, expression tissulaire, publications...). À titre d'exemple pour l'homme, la base de données UniProt référence 146000 protéines dont environ 21000 ont été revues par des experts.

Le NCBI (National Center for Biotechnology Information) est une base de données dédiée au référencement de l'information biologique multidisciplinaire (santé, biomédical, biologie végétale, biologie animale, transcriptomique, protéomique...). On qualifie le NCBI de « portail », car elle rassemble des données de tous types dans plusieurs bases de données indépendantes, des publications scientifiques (PubMed) et aussi des outils en ligne (comme le Blast par exemple ([Altschul et al., 1990](#))). Ainsi, le NCBI dispose de 66 bases de données indépendantes (comme GeneBank ([Benson et al., 2015](#)), HomoloGene ou Gene Expression Omnibus ([Barrett et al., 2013](#))), de 64 outils (comme le Blast et Genome Workbench (<http://www.ncbi.nlm.nih.gov/tools/gbench/>)) et un libre accès à 28 FTP (serveur de stockage de données brutes) parmi les 66 bases de données précédentes. Dans le cadre d'une étude de protéomique, il est possible d'interroger les bases NCBI Gene ou NCBI Protein de la même façon que pour UniProt. Ceci donne accès à une fiche unique composée de l'identifiant de la protéine, sa séquence, la date de sa découverte avec une publication associée et diverses informations sur des régions spécifiques de la séquence protéique. L'intérêt du NCBI est la mise en relation de toutes ces bases et outils. Ainsi, à partir de cette même fiche on a accès à la localisation chromosomique du gène codant la protéine (et à la cartographie du chromosome en question), aux protéines homologues dans d'autres espèces, à des outils d'alignement et d'analyses de la séquence protéique (recherche de protéines similaires ou de domaines par exemple).

L'EBI (European Bioinformatics Institute) est une plateforme scientifique européenne qui propose à la fois des outils et des bases de données biologiques, mais aussi des formations et des consortiums d'experts pour la validation des données biologiques. L'EBI dispose d'une plateforme unique dédiée aux outils et bases de données qui sont regroupés en 9 catégories : génétique et variation du génome, expression des gènes, protéomique, structures moléculaires et cellulaires, systèmes cellulaires (voies métaboliques et interactions moléculaires), biochimie et métabolomique, ontologies, bibliographie et des outils multidomains (workflows). L'EBI est une plateforme gratuite qui dispose d'outils populaires dans l'analyse de données « omiques » comme la base de données Ensembl ([Cunningham et al., 2015](#)), la base d'interactions moléculaire IntAct ([Orchard et al., 2014](#)) ou l'outil d'annotation par ontologie de gènes QuickGO ([Binns et al., 2009](#)).

KEGG (Kyoto Encyclopedia of Genes and Genomes) est une base de données cartographiant les génomes selon différentes catégories : les voies métaboliques, les orthologies d'espèce, les gènes et leur expression, les réactions biochimiques, les maladies humaines et les drogues. Elle dispose aussi d'outils liés à sa base de données comme l'annotation de séquences (via un blast local à KEGG). Toutes les données de cartographie de KEGG sont aussi visualisables grâce à une interface de visualisation (Figure 15).

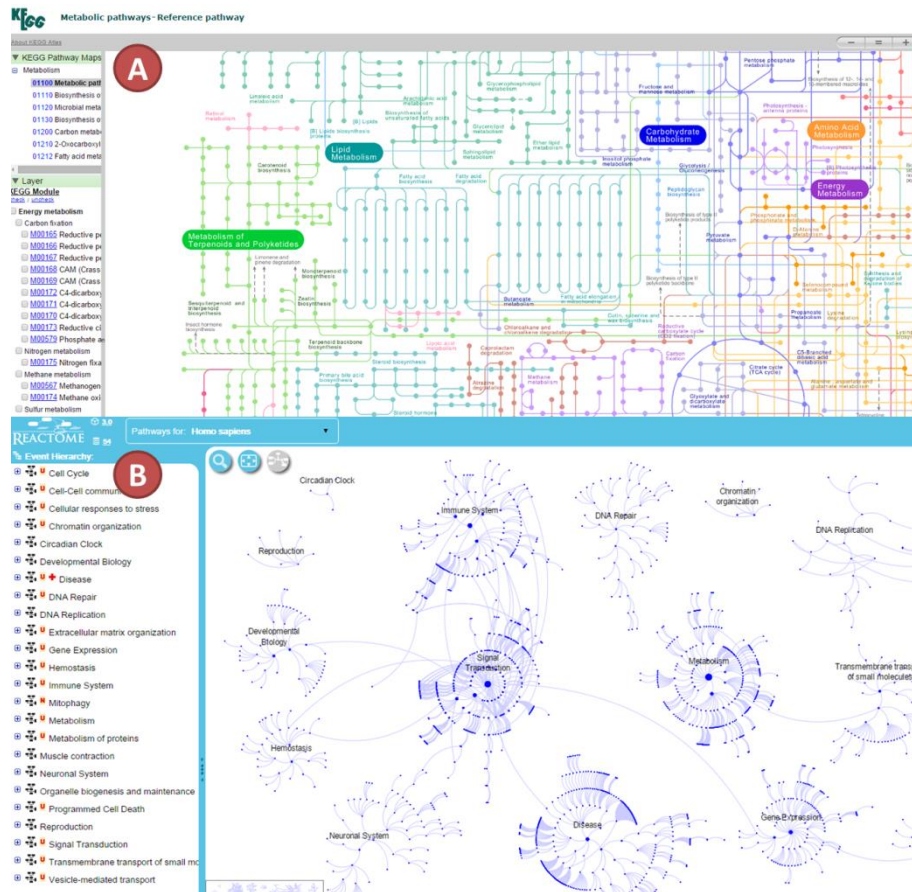


Figure 15 : (A) L'outil de visualisation KEGG Atlas est une interface graphique avancée pour explorer les cartographies de KEGG avec la capacité d'interagir avec les voies métaboliques pour en obtenir plus d'information. (B) L'outil de visualisation « Pathway Browser » de Reactome permet d'analyser des données et de cartographier les voies métaboliques ou d'enrichir les résultats avec d'autres voies disponibles dans la base.

Reactome est une base de données qui regroupe tous les événements biologiques au sein d'une même interface de visualisation. Chaque système biologique et voie métabolique est compartimenté dans des sous-ensembles uniques. La visualisation grâce à l'outil « Pathway Browser » permet d'interagir avec les voies métaboliques affichées et aussi de charger des informations sur ces voies (Figure 15). Reactome permet aussi d'analyser des données expérimentales pour les annoter avec sa base de données et les visualiser avec son interface, tout en gardant l'information sur les variations de l'expression des gènes par exemple.

GeneCards est une plateforme d'interrogation de bases de données humaines pour l'étude des gènes. Elle est divisée en plusieurs bases de données dédiées aux maladies, au développement embryonnaire et aux cellules souches (on y trouve notamment des données concernant le développement du muscle et du TA), aux voies métaboliques et à la localisation chromosomique des gènes. GeneCards est utile pour synthétiser les informations connues pour un gène et pour répertorier les outils biologiques disponibles grâce à une interface intuitive et simplifiée. La suite d'outils de GeneCards (nommée Gene Analytics) permet d'analyser des listes de gènes par la mise en relations de toutes ses bases.

Cas des ruminants

La base de données NCBI propose des sous-bases (grâce aux sections « Taxonomy » ou « Genome » du NCBI) dédiées à de nombreuses espèces et notamment aux ruminants comme la chèvre, le mouton ou le bovin. À ce jour chez le bovin, l'onglet « Taxonomy » du NCBI informe sur le nombre de protéines (environ 132000) et de gènes codant des protéines (environ 23500) identifiés, les données accessibles (8500 références de données de puces à ADN et plus de 3000 échantillons biologiques disponibles avec leurs méthodes de prélèvement) ou encore la description complète de chaque chromosome du génome. Il existe aussi des bases de données complètement dédiées à certaines espèces de ruminants : Bovine Genome Atlas pour le bovin ([Reese et al., 2010](#)), International Sheep Genomic Consortium pour le mouton ([Kijas et al., 2009](#)) et GoSh dédiée à la fois au mouton et à la chèvre ([Caprera et al., 2007](#)). Ces bases regroupent des informations sur les différentes races d'animaux, les protéines et gènes identifiés, des séquences et souvent quelques outils d'analyses de séquences (blasts, lecteur de génome et parfois annotation fonctionnelle). Il existe aussi des bases de données qui permettent de rassembler toutes les informations biologiques connues pour une liste de protéines ou de gènes chez les ruminants, c'est le cas d'AgBase (voir partie Gene Ontology ; ([McCarthy et al., 2011](#))) et de Bovinemine (<http://bovinegenome.org/bovinemine/>).

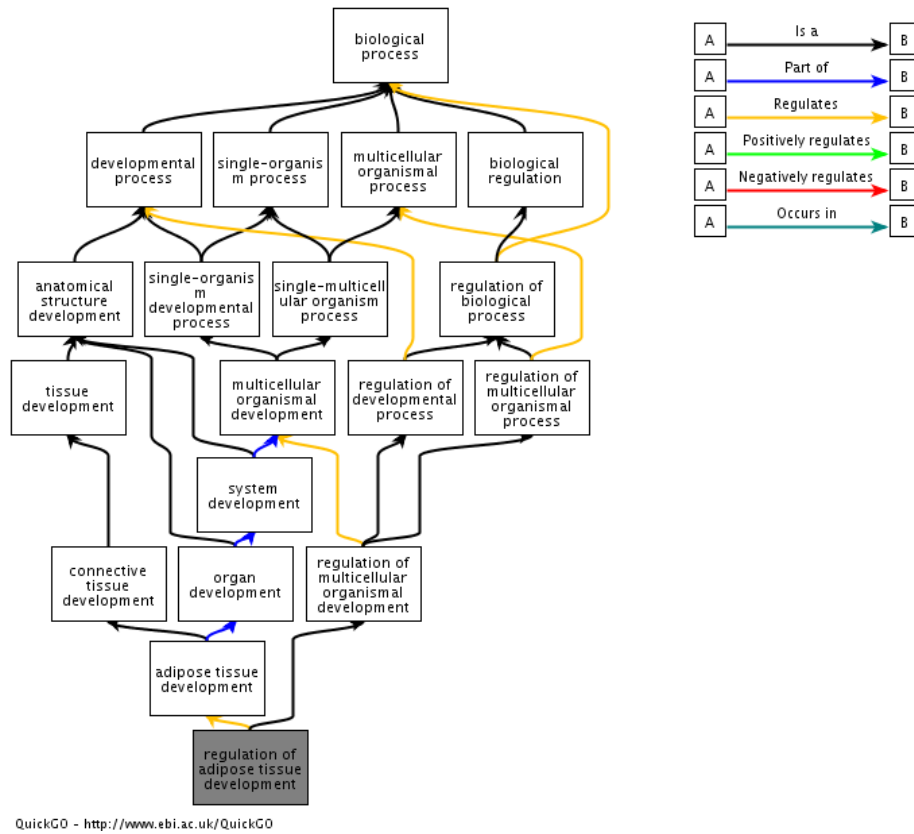


Figure 16 : Arbre représentatif des différents niveaux de relations amenant jusqu'au terme GO « regulation of adipose tissue development ». Les couleurs de flèche correspondent aux différentes relations possibles entre les termes GO (<http://www.ebi.ac.uk/QuickGO/>).

6.2 – Ontologie des gènes : l'annotation fonctionnelle des gènes et des protéines

Le projet d'ontologie des gènes, « Gene Ontology » (GO) en anglais, a été créé afin de structurer la description des gènes et des protéines dans le cadre d'une ontologie commune à toutes les espèces. En informatique et en science de l'information, l'ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations. C'est l'équipe du Gene Ontology Consortium qui a structuré et vulgarisé la GO (Ashburner *et al.*, 2000; Ben Jemaa *et al.*, 2013; Chokkalingam *et al.*, 2013) pour les gènes et les protéines. La GO est comparable à un dictionnaire hiérarchisé de termes GO liés entre eux par des liens de parenté (impliqués ou faisant partie d'un processus biologique) ou des relations fonctionnelles (régulation ou activation de processus biologiques ; Figure 16; (Huntley *et al.*, 2014)). Les protéines jouent un rôle dans une ou plusieurs dizaines de fonctions biologiques. C'est pourquoi on parle de protéines « annotées » par un ou plusieurs termes GO. Les termes GO sont des codes pour décrire trois types d'informations : les fonctions moléculaires, les processus biologiques et les composants/localisations cellulaires.

GO Evidence Code Decision Tree

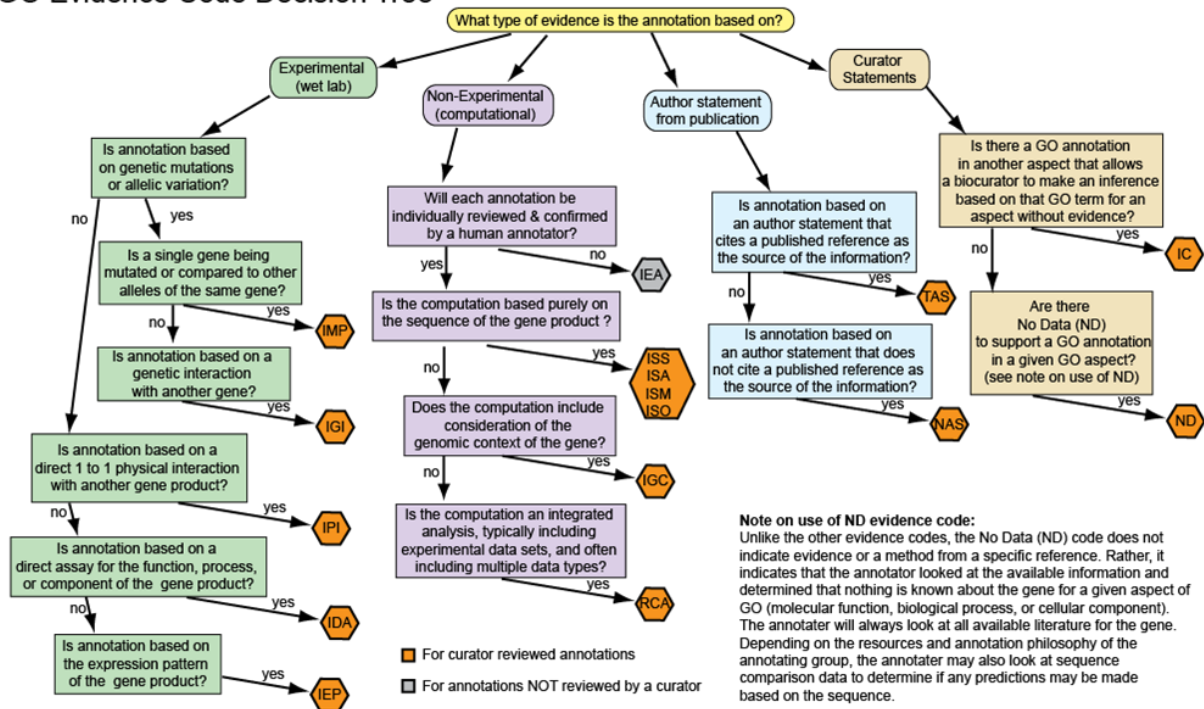


Figure 17: Arbre représentatif des différents niveaux de preuves des annotations par termes GO (<http://www.geneontology.org/>).

Pour un gène, l’annotation par des termes GO indique dans quels processus biologiques il est impliqué, ses différentes fonctions moléculaires et sa localisation cellulaire. Il existe différents types d’annotations : automatique (comparaison de séquences), manuelle (curation par des experts), fonctionnelle (prédiction de fonction biologique), relationnelle (famille de gènes, réseaux de régulation et d’interactions) et structurale (localisation et organisation des gènes sur le génome). Une fois annotée par un terme GO, une annotation est soumise à une notation par un code (nommé « Evidence codes » ; Figure 17) afin de préciser son niveau de preuve/fiabilité (annotation expérimentale avec résultats ou annotation prédite, par exemple). Les annotations électroniques (code IEA) représentent environ 90% des annotations totales (Skunca *et al.*, 2012). Chaque terme GO est composé d’une fiche descriptive et de la liste des protéines et gènes qu’il annote (Figure 18). On recense actuellement plus de 23 millions de gènes et de protéines annotés pour 42,5 millions d’annotations par des termes GO (Huntley *et al.*, 2014). Chez l’homme, on répertorie environ 479000 annotations pour 47000 protéines/gènes annotés (source AmiGO). Ces annotations sont spécifiques à une espèce ou communes à plusieurs. Une protéine (ou un gène) peut être annotée par plusieurs termes GO spécifiques. À noter que dans la plupart des espèces, de nombreuses protéines n’ont pas encore de fonction connue et sont donc peu ou pas annotées (Hanson *et al.*, 2010).

glycolytic process A

Term Information

Accession GO:0006096
Name glycolytic process
Ontology biological_process
Synonyms alt. id: GO:0019641
alt. id: GO:0019642
Emden-Meyerhof pathway
Emden-Meyerhof-Parnas pathway
anaerobic glycolysis
glycolysis
modified Emden-Meyerhof pathway

Definition The chemical reactions and pathways resulting in the breakdown of a carbohydrate into pyruvate, with the concomitant production of a small amount of ATP and the reduction of NAD(P) to NAD(P)H. Glycolysis begins with the metabolism of a carbohydrate to generate products that can enter the pathway and ends with the production of pyruvate. Pyruvate may be converted to acetyl-coenzyme A, ethanol, lactate, or other small molecules. *Source:* GOC:bf, [Wikipedia:Glycolysis](#), ISBN:0716720094, GOC:dph, ISBN:0879010479, ISBN:0201090910

Comment None
History See term [history](#) for GO:0006096 at QuickGO
Subset gosubset_prok

Community **GN** Add usage comments for this term on the GONUTS wiki.
[View](#) and [edit](#) information about this term on Wikipedia.

Related
[Link](#) to all **genes and gene products** associated to glycolytic process.
[Link](#) to all direct and indirect **annotations** to glycolytic process.
[Link](#) to all direct and indirect **annotations download** (limited to first 10,000) for glycolytic process.

Feedback Contact the [GO Helpdesk](#) if you find mistakes or have concerns about the data you find here.

Free-text filtering X

Your search is pinned to these filters
+ document_category: annotation
+ regulates_closure: GO:0006096
No current user filters.
▶ [Source](#)

Found entities B

Total: 2576; showing 1-10 Results count

<input type="checkbox"/>	Gene/product	Gene/product name	Qualifier	Direct annotation	Annotation extension	Assigned by	Taxon	Evidence	Ev
<input type="checkbox"/>	BPGM	Bisphosphoglycerate mutase		glycolytic process		UniProt	Bos taurus	IEA	UniProtKB-K
<input type="checkbox"/>	ADPGK	ADP-dependent		glycolytic		UniProt	Bos	IEA	UniPathway

Figure 18 : Fiche descriptive du terme GO « glycolytic process » (glycolyse) provenant de la base de données AmiGO. La partie A synthétise les informations liées au terme GO (code, dénomination, définition, etc...). La partie B représente toutes les entités annotées par le terme GO (protéines, gènes, « Evidence code » de l'annotation, espèce, base de données source de l'information...).

Utilisation avancée de la Gene Ontology

L'annotation d'une liste d'identifiants est réalisable en interrogeant des bases de données de termes GO (protéine par protéine ; travail souvent long et fastidieux) ou via des outils en ligne (analysant l'intégralité du jeu de données et proposant un tri ou une hiérarchisation des résultats). De nombreux outils et bases de données sont créés chaque année avec de nouvelles méthodes de calculs, des données mieux vérifiées ou spécifiques à une espèce ([Dameron et al., 2013](#)). Ils peuvent être spécifiques à une espèce ou uniquement composés de données vérifiées par des experts. Les bases de données les plus couramment utilisées sont AmiGO ([Carbon et al., 2009](#); [Gene Ontology, 2015](#)), UniProt-GOA ([Huntley et al., 2015](#)) et QuickGO ([Binns et al., 2009](#)). Elles sont mises à jour quotidiennement et permettent de rechercher des informations par terme GO ou par identifiant de protéine ou de gène. Pour identifier les fonctions biologiques majeures dans lesquelles sont impliquées les protéines ou les gènes d'une liste, on utilise des outils qui annotent et qui répertorient tous les termes GO pour une liste donnée. Les outils en ligne les plus couramment utilisés sont DAVID ([Huang et al., 2009](#)), GoMiner ([Verovskaya et al., 2013](#)), FatiGO ([Al-Shahrour et al., 2004](#)) et GOrilla ([Eden et al., 2009](#)). Ces outils s'accompagnent souvent de tests statistiques afin de mettre en évidence les termes GO les plus représentatifs du jeu de données, on parle alors d'enrichissement. Certains outils possèdent leurs propres bases de données de termes GO provenant souvent des

bases citées précédemment ayant été filtrées par une curation spécifique à l'outil. Il existe aussi des outils pour la GO à intégrer dans des scripts de programmation comme : le package Perl « GO::TermFinder » ([Rivenbark et al., 2013](#)), le package R « GO.db » référencé sur Bioconductor (<http://www.bioconductor.org/packages/2.13/data/annotation/html/GO.db.html>; ([Lawrence et al., 2013](#))), ou le package « BinGO » ([Maere et al., 2005](#)) du logiciel de visualisation de réseaux Cytoscape (partie 6.4 ; ([Smoot et al., 2011](#))). Ces outils permettent une plus grande flexibilité de recherche et d'affichage des résultats, mais nécessitent des compétences en programmation informatique.

La GO est une étape essentielle pour analyser un jeu de données de génomique. L'objectif est de mettre en évidence les fonctions biologiques les plus représentatives d'un jeu de données grâce à des méthodologies de calcul d'enrichissement des résultats d'annotation. Par exemple, pour une liste de protéines identifiées expérimentalement, l'enrichissement a pour objectif de révéler les termes GO surreprésentés. Il existe plusieurs outils qui réalisent un enrichissement de termes GO avec une liste de protéines (ou de gènes) : l'outil de classification PANTHER ([Mi et al., 2013](#)), notamment utilisé par le GO Consortium, l'outil FatiGO+ ([Al-Shahrour et al., 2007](#)) ou encore l'outil Ontologizer ([Bauer et al., 2008](#)). Ces outils utilisent des méthodes de calcul qui leur sont propres pour classer les termes GO, mais ces méthodes sont basées sur des comparaisons 2 à 2 d'échantillon grâce à un test exact de Fisher. L'outil g:Profiler utilise sa propre méthode d'enrichissement nommée g:SCS qui utilise des tables de contingence 2 à 2 mais prend aussi en compte l'annotation effectuée par rapport à l'arborescence complète de la GO ([Reimand et al., 2007](#)). D'autres outils, comme GO:Stat ([Beissbarth et Speed, 2004](#)), utilisent le test de Khi2 pour calculer l'enrichissement d'un terme GO dans le cas de grandes listes de gènes. De manière générale, le calcul d'enrichissement d'un terme GO se fait avec un test exact de Fisher (Tableau 5). Les résultats obtenus se présentent sous la forme d'une liste classée par ordre croissant de la valeur de significativité des termes GO (ou p -value). Plus la p -value est faible, plus le terme GO est dit « enrichi » et plus la présence de ce terme dans la liste de gènes étudiés est significative. La p -value d'un terme GO représente la probabilité d'obtenir ce terme dans les résultats d'annotation si on effectuait un tirage au hasard. Souvent un seuil est choisi pour déterminer si l'enrichissement du terme est suffisant pour être gardé dans les résultats significatifs. On associe couramment un ajustement de la p -value par une correction comme le test de Benjamini & Hochberg ([Benjamini et Hochberg, 1995](#)) moins strict que le test de Bonferroni. Cette correction est nécessaire pour pallier au problème du taux de fausse découverte (FDR en anglais). Le FDR représente la proportion de résultats faux parmi tous les résultats. Dans notre cas, le FDR représente le nombre de fois qu'un terme GO n'est pas enrichi parmi tous les résultats du test exact de Fischer. Appliquer une telle correction permet de vérifier si un terme GO n'est pas déclaré enrichi à tort. C'est souvent les termes GO les plus généraux de l'ontologie qui sont parmi les plus « enrichis », car ils annotent plus de protéines et recouvrent plus aisément un jeu de données. Cependant, dans certaines études, les biologistes recherchent des fonctions plus précises, souvent uniques à quelques protéines et issues de successions de fonctions biologiques plus générales. Ainsi, ces fonctions précises se retrouvent avec des p -values élevées, car moins priorisées par l'enrichissement, mais sont tout de même à considérer.

Table de contingence 2 x 2	Nombre d'identifiants dans la liste de référence	Nombre d'identifiants dans la liste expérimentale	Totaux
Annotés par le terme GO	X	Z	$\sum_{1,*}$
Non annotés par le terme GO	N - X	T - Z	$\sum_{2,*}$
Totaux	$\sum_{*,1}$	$\sum_{*,2}$	$\sum_{*,*}$

Tableau 5 : Table de contingence deux à deux utilisée pour calculer l'enrichissement d'un terme GO qui annote une liste Z de gènes par rapport à la liste X des gènes de l'espèce.

Cas des ruminants

En novembre 2015, on répertorie pour le bovin environ 203000 annotations GO pour 22000 protéines/gènes annotés, pour la chèvre 103000 annotations pour 13000 protéines/gènes annotés et pour le mouton environ 3900 annotations pour 350 protéines/gènes annotés (source AmiGO). La base de données nommée AgBase (McCarthy et al., 2011) est composée de termes GO dédiés aux espèces agricoles, dont certains ruminants (comme le bovin et le mouton), aux animaux domestiques et aux végétaux de culture. Les données d'AgBase sont vérifiées par les curateurs internes et proviennent d'autres bases de données ou sont spécifiques à cette base. Dans AgBase, on répertorie chez le bovin 3 fois plus d'annotations et de protéines/gènes annotés que dans AmiGO. Cependant 75% de ces annotations sont électroniques et non vérifiées par des experts (<http://www.agbase.msstate.edu/cgi-bin/tools/goprofiler.pl?db=Bovine>). Un autre outil d'analyse de terme GO nommé agriGO (Du et al., 2010) est dédié aux animaux d'élevage (dont le bovin et la chèvre) et à de nombreux végétaux de culture. Cet outil rassemble les termes GO annotant les protéines ou les gènes d'une liste sous la forme d'un tableau. On y visualise les termes GO les plus représentatifs (« enrichis ») et les protéines ou gènes qu'ils annotent. Deux graphiques sont produits pour compter le nombre de protéines ou de gènes annotés par terme GO, ainsi qu'un arbre de parenté entre les termes GO trouvés dans le jeu de données (Figure 19). Quel que soit l'outil, les ruminants restent mal annotés, de plus ces annotations sont souvent électroniques et non vérifiées. Il existe finalement peu d'outils ou bases de données dédiés à la GO des identifiants de ruminants.

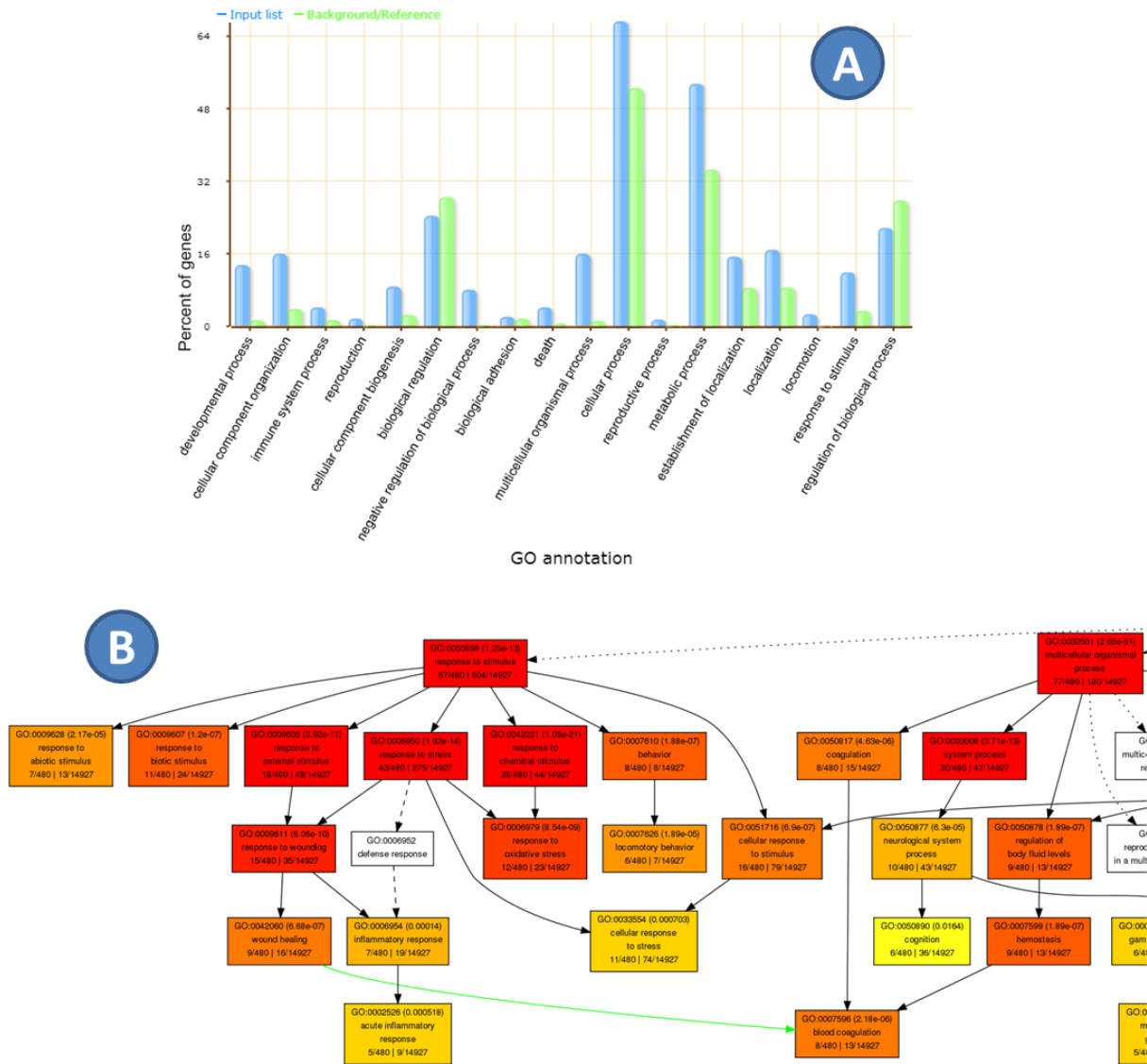


Figure 19 : Résultats graphiques obtenus en utilisant l’outil agriGO. (A) Le graphique en barres représente les termes GO les plus généraux retrouvés dans l’analyse en fonction du pourcentage de gènes annotés (les barres bleues donnent accès aux gènes annotés et les barres vertes représentent les annotations du terme GO selon l’espèce d’étude). (B) agriGO représente aussi les résultats sous la forme d’un arbre de parenté, la couleur des termes représente les degrés d’enrichissement des termes (rouge pour « très enrichis » à blanc pour « non enrichis » ; les carrés donnent aussi accès aux gènes annotés).

6.3 – La prédiction des protéines sécrétées

L’étude du sécrétome

Le sécrétome rassemble les protéines sécrétées par une cellule. Les protéines sont sécrétées soit entre deux cellules proches (sécrétion paracrine), soit vers le sang pour atteindre une autre cellule ou un autre tissu (sécrétion endocrine). Les protéines sécrétées représentent environ 10% du génome d’une espèce (Pavlou et Diamandis, 2010; Skalnikova et al., 2011). Chez les eucaryotes, au moins cinq voies de sécrétion des protéines ont été mises en évidence. La voie de sécrétion classique des protéines néo-synthétisées est leur translocation dans la lumière du réticulum endoplasmique

puis leur transport dans l'appareil de Golgi et leur sécrétion par exocytose. Cette voie de sécrétion concerne les protéines qui possèdent un peptide signal N-terminal. Les quatre autres voies de sécrétion « non-classiques » concernent les protéines sans peptide signal et qui ne sont donc pas dépendantes de l'appareil de Golgi et du réticulum endoplasmique. Ces autres voies de sécrétion impliquent le recyclage de l'endosome, des transporteurs membranaires, des translocations de protéines des feuillettes internes aux feuillettes externes de la membrane plasmique (membrane flip-flop), et le bourgeonnement de la membrane qui conduit à la formation de vésicules ou des exosomes contenant entre autre des protéines ([Nickel, 2003](#)).

Des études expérimentales du sécrétome sont conduites, mais présentent des inconvénients à considérer : la présence de protéines intracellulaires retrouvées dans le compartiment extracellulaire (après rupture des cellules par exemple lors de la préparation des cellules ou associées aux débris cellulaires insolubles), la faible concentration des protéines sécrétées qui nécessite des méthodes d'enrichissement pour pouvoir les identifier, la large fourchette de concentrations (10 à 13 log) des protéines notamment pour le plasma qui ne permet qu'une identification partielle du protéome ([Skalnikova et al., 2011](#); [Weigert et al., 2014](#)). Aussi, des méthodes de prédiction des protéines sécrétées ont été développées ([Klee et Ellis, 2005](#)). Les données nécessaires à la prédiction *in silico* de peptides signaux sont les séquences protéiques.

De nombreux outils ont pour objectifs de prédire les protéines potentiellement sécrétées grâce à un peptide signal et des bases de données répertorient ces protéines sécrétées ([Caccia et al., 2013](#)). Les protéines sécrétées grâce à un peptide signal sont les plus couramment recherchées en protéomique, car le peptide signal est détecté directement par une analyse de la séquence de la protéine. Des techniques d'alignement évaluent la présence d'un motif « signal » sur une séquence protéique avec des scores de précision raisonnables ([Caccia et al., 2013](#); [Emanuelsson et al., 2007](#)). Les outils les plus communément utilisés (Tableau 6) sont SignalP ([Petersen et al., 2011](#)), Phobius ([Kall et al., 2007](#)), SigCleave ([Rice et al., 2000](#)) et PrediSi ([Hiller et al., 2004](#)). Klee et al. ([Klee et Ellis, 2005](#)) ont ainsi prédit le sécrétome putatif humain, porcine et du poisson Fugu en associant une comparaison de séquences BLAST à l'identification des peptides signaux ([Klee et al., 2004](#)).

Pour les voies de sécrétion qui n'impliquent pas un peptide signal, il est plus difficile de prédire une protéine sécrétée grâce à sa séquence. D'autres outils prédisent la localisation cellulaire de la protéine par analyse de la séquence, par exemple : TargetP ([Emanuelsson et al., 2000](#)), PSORTb ([Yu et al., 2010](#)), SecretomeP ([Bendtsen et al., 2004](#)) ou Cell-PLoc ([Chou et Shen, 2008](#)). Le résultat est la localisation de la protéine dans la cellule. Cette localisation n'est pas forcément signe de sécrétion de la protéine, mais permet d'émettre une hypothèse vérifiable à l'aide de la littérature. Pour améliorer ces prédictions, certaines bases de données associent la prédiction d'un peptide signal et la localisation cellulaire, c'est le cas de la base UniProt (souvent un article concernant la sécrétion de la protéine y est associé). La base de données Pfam (partie 6.1) propose en plus de l'analyse de la séquence, une multitude de fiches descriptives de protéines selon les domaines sur la séquence. On y retrouve les protéines avec un signal peptide mis en évidence expérimentalement (citation directe de l'article), ou prédit par l'outil SignalP ([Nielsen et al., 1999](#)). La base de données Peptide Atlas ([Deutsch, 2010](#)) référence un maximum de peptides pour de nombreuses espèces, on peut notamment y trouver un certain nombre de peptides signaux pour les espèces modèles comme l'homme, le rat ou la drosophile. La base LOCATE rassemble les données de localisation cellulaire des protéines identifiées chez l'homme et de la souris ([Fink et al., 2006](#); [Sprenger et al., 2008](#)). Les

emplacements subcellulaires ont été déterminés expérimentalement par immunofluorescence à haut débit et par une analyse informatique (text mining) des connaissances disponibles et publiées pour chaque protéine. Elle propose ainsi plus de 4000 protéines sécrétées chez l'homme et la souris. La base de données Secreted Proteins Database ([Chen et al., 2005](#)) référence les protéines sécrétées chez l'homme, le rat et la souris, mais la base de données n'est malheureusement plus mise à jour depuis 2008. Actuellement, la prédiction d'une forme de sécrétion pour une protéine basée sur un outil n'est plus suffisante à cause notamment des taux d'erreurs. Afin d'appuyer les prédictions des outils, les chercheurs commencent à combiner d'autres informations telles que la localisation cellulaire obtenue par GO comme preuve supplémentaire ([Huang, 2012](#)) à la prédiction et à la localisation extracellulaire des protéines.

Outils	Type de prédiction	Taux de faux positif (%)	Fichier d'entrée	Site de référence
<i>Prédiction de peptides signaux</i>				
SignalP	Prédiction de signal peptide	3	Séquence FASTA (max 2000)	http://www.cbs.dtu.dk/services/SignalP/
PrediSi	Prédiction de signal peptide	12	Séquence FASTA (une ou plus à la fois)	http://www.predisi.de/
SigCleave	Prédiction de signal peptide	5	Séquence FASTA (une ou plus à la fois)	http://emboss.bioinformatics.nl/cgi-bin/emboss/sigcleave
Phobius	Prédiction de signal peptide	16	Séquence FASTA (une ou plus à la fois)	http://phobius.sbc.su.se/
<i>Autres voies de sécrétion</i>				
TargetP	Prédiction de localisation cellulaire	10	Séquence FASTA (moins de 4000aa et 2000 max)	http://www.cbs.dtu.dk/services/TargetP/
PSORTb	Prédiction de localisation cellulaire	0 (prédiction contre des motifs)	Séquence FASTA (100 max)	http://www.psort.org/psortb/
SecretomeP	Prédiction de protéine sécrétée par une autre voie	5	Séquence FASTA (moins de 4000aa et 2000 max)	http://www.cbs.dtu.dk/services/SecretomeP/

Tableau 6 : Les outils les plus communément utilisés pour la prédiction de protéines potentiellement sécrétées grâce à un peptide signal ou par d'autres voies de sécrétion (prédiction de la localisation cellulaire). Pour chaque outil, les auteurs ont calculé les taux de faux positifs liés à leur analyse qui sont reportés ici.

Cas des ruminants

Actuellement, il n'existe pas de base de données dédiée aux protéines sécrétées chez les ruminants. Aucun outil de prédiction de protéines sécrétées ou d'aide à la recherche de protéines potentiellement sécrétées n'est dédié aux ruminants. Cependant, la plupart des outils de prédiction utilisent des séquences au format FASTA pour rechercher les motifs de peptides signaux et ceux-ci fonctionnent pour toutes les espèces. La base de données UniProt, rapporte environ 675 protéines considérées comme sécrétées chez le bovin comparativement aux 1800 annotées chez l'homme (validées par experts ou électroniquement par comparaison de séquence ; détectées via les protéomes de référence de la base UniProt). Dans l'état actuel des connaissances, aucune étude n'a été réalisée sur le sécrétome bovin du TA et du muscle. On référence néanmoins une étude expérimentale sur le sécrétome du plasma bovin ([Sun et al., 2013](#)). La plupart des protéines indiquées comme sécrétées ont été annotées par similarité de séquence entre espèce ou manuellement par des experts. Il existe quelques protéines, étudiées indépendamment, dont la localisation cellulaire a été expérimentalement validée (seulement 3% des protéines indiquées comme sécrétées par UniProt) : c'est le cas de la Serpine A3-1 ([Gagaoua et al., 2015](#)), de l'Interleukine-21 ([Muneta et al., 2003](#)) et de IGIP (IgA-inducing protein ; [Austin et al., 2003](#)).

6.4 – Les interactions de protéines et voies métaboliques

Les interactions protéine-protéine

Une interaction protéine-protéine (PPI) apparaît lorsque deux ou plusieurs protéines se lient entre elles pour réaliser une fonction biologique (régulation d'une voie métabolique par exemple). Une protéine peut interagir avec une autre protéine, ou avec un acide nucléique (ADN – peptide), ou encore avec des petites molécules (liaison avec des drogues ou des lipides par exemple ; ([Hermjakob et al., 2004](#); [Jin et al., 2014](#))). Les interactions moléculaires sont qualifiées en fonction du type d'interactions et des méthodes qui ont permis leur détection ([Figure 20](#) ; [Rao et al., 2014](#))). Il existe de nombreuses méthodes de détection qui peuvent être expérimentales ou prédictives ([Carneiro et al., 2015](#)). La base de données Ontology Lookup Service ([Cote et al., 2010](#)) permet de visualiser les différentes qualifications possibles pour une interaction (ou une annotation fonctionnelle par des termes GO) et ainsi de trier les interactions en fonction des critères souhaités (interaction revue par expert, méthode de détection expérimentale...).

Plus d'un millier de méthodes pour détecter une interaction moléculaire ont été recensées ([Cote et al., 2010](#)) ; <http://psidev.cvs.sourceforge.net/viewvc/psidev/psi/mi/rel25/data/psi-mi25.obo>). Il existe les méthodes expérimentales biologiques et chimiques par exemple : le double hybride (cette méthode classique utilise l'activité de transcription en tant que mesure d'une interaction protéine-protéine.), la polymérisation (interaction déduite en mesurant la polymérisation / dépolymérisation d'un interacteur), les études enzymatiques (enzyme ou substrat participant à une réaction biochimique) ou des méthodes de chromatographie (comme la coimmunoprécipitation ; dans cette méthode, un anticorps spécifique à la molécule d'intérêt est utilisé pour la séparer de son mélange moléculaire et capturer son ligand simultanément). Il y a des méthodes expérimentales de physiques et d'imagerie comme : la résonance électromagnétique (une forme de spectroscopie dans laquelle l'absorption des micro-ondes par un échantillon dans un fort champ magnétique est

utilisée pour étudier les atomes ou les molécules avec des électrons non appariés) ou par microscopie fluorescente (cette microscopie utilise une lumière de haute intensité pour illuminer l'échantillon, des zones de l'échantillon émettent alors une fluorescence). Il y a aussi des méthodes expérimentales directes comme l'inférence génétique (affecter l'activité d'un gène spécifique), les modifications post-transcriptionnelles ou la fragmentation de protéines (deux protéines sont liées entre elles par un fragment de protéine support, puis introduites dans un milieu de culture pour vérifier leur potentielle interaction par fluorescence). Il existe des méthodes de prédiction d'interactions dans le cas où les méthodes biologiques ne sont pas disponibles ou sont irréalisables. Des méthodes prédictives manuelles effectuées par des experts curateurs et par homologie entre espèces (il ne s'agit là que de déductions). Il y a aussi les méthodes prédictives informatiques comme la prédiction de structures protéiques (on parle de « docking », prédiction de liaison entre des structures de protéines) ou la prédiction de l'interaction en fonction des connaissances (avec du « text mining », fouille dans la littérature de rapprochement entre deux protéines par des intermédiaires ou directement entre elles).

Certaines bases de données de PPI ont mis en place des systèmes de curations basés sur un partenariat entre les bases et des règles strictes. Deux systèmes de curation ont été créés pour rendre les données de PPI plus homogènes entre toutes les bases disponibles. Le projet IMEx (International Molecular Exchange consortium ([Orchard et al., 2012](#))), qui rassemble 11 bases de données, souhaite partager les expertises de revue des interactions afin d'éviter les redondances et de proposer des interactomes validés par des experts. Le projet MIMIx (The Minimum Information about a Molecular Interaction eXperiment ([Orchard et al., 2007](#))), dédié à la structuration des données d'interactions moléculaires par une meilleure définition d'une interaction et des méthodes utilisées pour la mettre en évidence. Une PPI doit être prouvée expérimentalement pour pouvoir appartenir au projet IMEx ou MIMIx. Les interactions moléculaires sont de plus en plus étudiées, car elles permettent de comprendre rapidement les liens fonctionnels entre des molécules ([Biron et Bedard, 2015](#); [Carneiro et al., 2015](#); [Sable et Jois, 2015](#)). De ce fait, de nombreuses bases de données dédiées à une espèce spécifique ou à un type d'interaction voient le jour et les bases les plus populaires se rassemblent en consortium pour unifier les ressources disponibles ([Klingstrom et Plewczynski, 2011](#)).

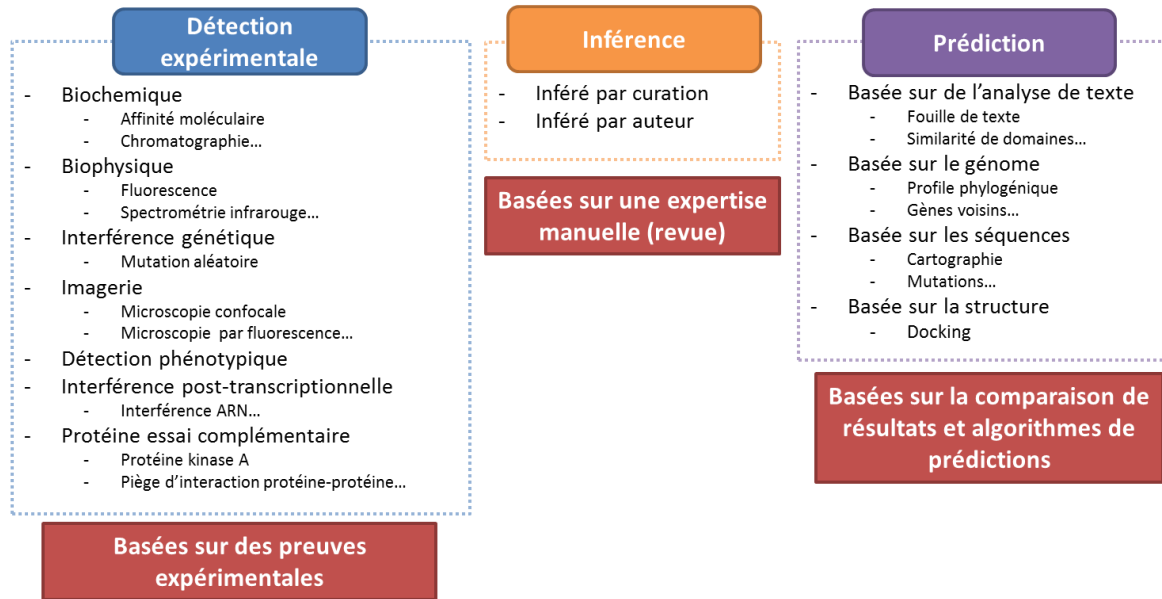


Figure 20 : Représentation non exhaustive des méthodes de détection des interactions moléculaires répertoriées sur le site OLS (Ontology Lookup Service ([Cote et al., 2010](#))).

Les bases de données d'interactions de protéines

Il existe une multitude de bases de données permettant de rechercher des interactions pour une liste de protéines ou de gènes. Elles peuvent être séparées en bases de données indépendantes ou rassemblées et triées dans une méta-base de données.

Les bases indépendantes IntAct ([Orchard et al., 2014](#)), BioGRID ([Chatr-Aryamontri et al., 2015](#)), MINT ([Licata et al., 2012](#)), BIND ([Isserlin et al., 2011](#)), DIP ([Salwinski et al., 2004](#)) et HPRD ([Keshava Prasad et al., 2009](#)) sont des bases de données sources de PPI. Elles n'incorporent pas de données provenant d'autres bases ([Figure 21](#) et [Tableau 7](#)). Le nombre de PPI qu'elles recensent est très variable ([Tableau 6](#)). Ces bases sont multi-espèces, multi-interactions, avec un système de curation lié au moins au projet IMEx ou MIMIx et une validation expérimentale décrite dans au moins 2 articles de 2 journaux différents. Les PPI doublons entre ces bases de données sources sont rares et assurent une faible redondance ([Kalathur et al., 2014](#); [Prieto et De Las Rivas, 2006](#)). Les bases de données telles que DroID (PPI chez la drosophile ; [Murali et al., 2011](#)), MatrixDB (PPI extracellulaire ; [Launay et al., 2015](#)), InnateDB (PPI du système immunitaire ; [Breuer et al., 2013](#)) et MPIDB (PPI des microbes ; [Goll et al., 2008](#)) combinent la fouille de données disponibles dans d'autres bases de données de PPI avec leurs propres curations par experts. La combinaison de la fouille de données de grandes bases de PPI avec des curations par experts génère des bases de PPI plus fiables et attractives pour les chercheurs biologistes. Des recherches dans plusieurs bases de données sont donc complémentaires dans le sens où chaque nouvelle recherche donne accès à un grand ensemble d'interactions qui ne sont pas forcément référencées partout. De plus, [Martha et al. \(2011\)](#) ont mis en évidence que la fiabilité des résultats de recherche de PPI est augmentée lorsque les PPI sont déclarées dans au moins 2 bases de données différentes.

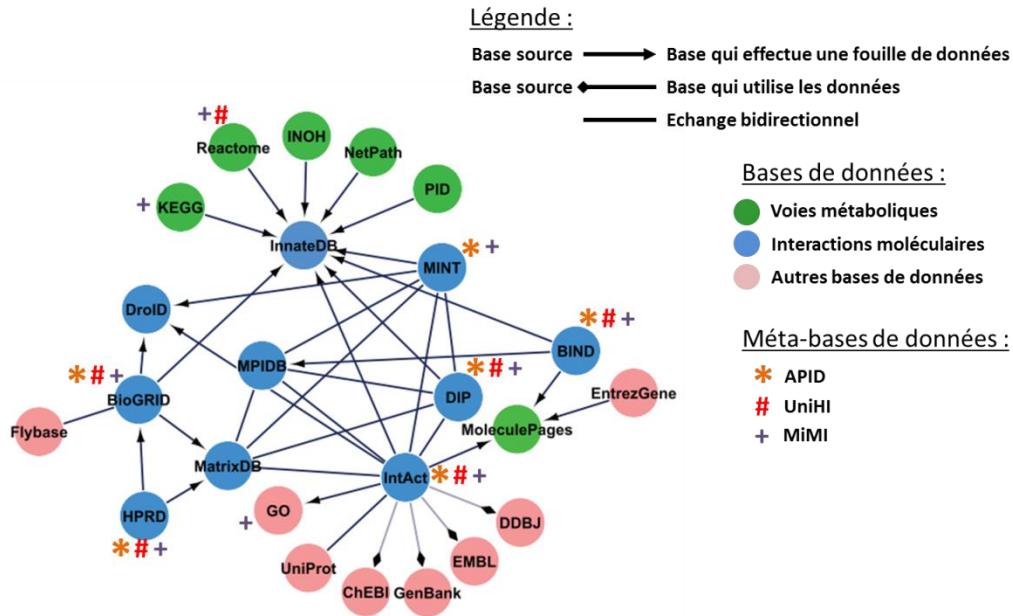


Figure 21 : Liens entre les bases de données d’interactions moléculaires (modifié à partir de (Klingstrom et Plewczynski, 2011)) et leurs appartenances à des méta-bases de données d’interactions.

Base de données	Nombre d'interactions	Type d'interactions	Site internet
BIND	190 000	PPi NPi SPi	https://www.bindingdb.org/
BioGrid	750 000	PPi CURE	http://www.thebiogrid.org/
DIP	110 000	PPi CURE IMEx MIMIx	http://dip.doe-mbi.ucla.edu/dip/Main.cgi
Droid	510 000	PPi NPi	http://www.droidb.org/
HPRD	41 000	PPi CURE	http://www.hprd.org/
InnateDB	32 000	PPi NPi CURE MIMIx	http://www.innatedb.com/
IntAct	380 000	PPi NPi SPi CURE IMEx MIMIx	http://www.ebi.ac.uk/intact/
MatrixDB	25 000	PPi SPi CURE IMEx	http://matrixdb.ibcp.fr/
MINT	130 000	PPi CURE IMEx MIMIx	http://mint.bio.uniroma2.it/mint/Welcome.do
Reactome	140 000	PPi MIMIx PRED	http://www.reactome.org/

Tableau 7 : Listes des principales bases de données d’interactions moléculaires représentées dans la Figure 21. Sont rapportés les liens vers les sites internet des bases, le nombre d’interactions référencées en octobre 2015 et aussi les types d’interactions : PPi (protéine-protéine interaction), NPi (acide-nucléique-protéine interaction), SPi (petite molécule-protéine interaction), CURE (curation interne des interactions), IMEx et MIMIx (pour l’appartenance à un consortium de PPi), et PRED (prédiction d’interactions).

APID ([Prieto et De Las Rivas, 2006](#)), MiMI ([Tarcea et al., 2009](#)) et UniHI ([Kalathur et al., 2014](#)) sont des méta-bases de données : elles rassemblent les PPI d'autres bases dans une seule source méta-base de données de PPI ([Figure 21](#)). APID référence uniquement des données expérimentales à partir des bases de données d'interactions et intègre les données de BioGRID, BIND, DIP, HPRD, IntAct et MINT. UniHI intègre les PPI identifiées expérimentalement uniquement chez l'homme. En plus des bases BioGRID, BIND, DIP, HPRD et IntAct, UniHI comprend aussi les voies métaboliques de Reactome ([Croft et al., 2014](#)), les PPI extraites par « text mining » dans les résumés Medline et les PPI orthologues dans des espèces modèles animales (prédiction d'interaction chez l'homme ([Brown et Jurisica, 2005](#)), HomoMint ([Persico et al., 2005](#)) et OrthoDB ([Kriventseva et al., 2015](#))). La méta-base MiMI n'est plus fonctionnelle, elle intégrait les données de BioGRID, BIND, DIP, HPRD, IntAct et MINT. Elle répertoriait également des interactions issues de KEGG, des voies métaboliques de Reactome et également des informations de Gene Ontology.

Toutes les bases de données citées précédemment peuvent être interrogées indépendamment en ligne ou via des outils qu'elles proposent. Il existe aussi un service en ligne, PSICQUIC ([Aranda et al., 2011](#)), qui permet d'interroger 32 bases d'interactions moléculaires en une seule requête. Pour une protéine unique, PSICQUIC liste toutes les interactions connues pour cette protéine parmi les 125 millions disponibles ([del-Toro et al., 2013](#)). Les interactions obtenues sont rangées dans un tableau avec les identifiants des protéines qui interagissent, une publication référençant l'interaction, la méthode de détection et, selon les bases, des indices de confiance. PSICQUIC permet aussi de visualiser rapidement les résultats sous forme d'un réseau. Cependant, PSICQUIC ne permet pas d'interroger en ligne des listes de protéines ou de gènes. PSICQUIC autorise cependant les interrogations multiples via son web service, il est dans ce cas nécessaire d'écrire son propre programme d'interrogation.

Réseaux d'interactions de protéines

Les bases de données référençant les interactions moléculaires présentent presque systématiquement leurs résultats sous forme de tableaux. Cependant, il est plus facile de visualiser puis d'analyser, ces résultats sous forme d'un réseau d'interactions avec des outils « topographiques » simples ou complexes ([Agapito et al., 2013](#)). Techniquement, un réseau de PPI est une représentation visuelle suivant la théorie des graphes ([Berge, 1958](#)) où les « nœuds » représentent les protéines et les « traits » représentent les liens d'interaction biologiques entre ces protéines (activation, régulation, interaction directe...). Il s'agit donc de graphes construits à partir de données sur les interactions par paires de protéines. Ces réseaux sont construits pour des paires de protéines qui appartiennent à un jeu de données. De plus, afin d'identifier de nouvelles protéines jouant un rôle dans une fonction biologique d'intérêt, des protéines d'un jeu de données peuvent être appariées avec d'autres protéines non présentes dans ce jeu de données, mais renseignées comme des interacteurs potentiels dans des bases de données ([Lievens et al., 2010](#)).

Des outils permettent la recherche d'interactions pour des listes de protéines et leur représentation sous forme de réseaux. C'est notamment le cas de string : String ([Franceschini et al., 2013](#)), BIPS (spécifique à l'Homme ; ([Garcia-Garcia et al., 2012](#))) ou encore DroPNet (spécifique à la drosophile ; ([Renaud et al., 2012](#))). La représentation des interactions moléculaires s'est standardisée par la mise en place des réseaux moléculaires, ceux-ci permettant de visualiser directement les voies

métaboliques et les cascades de régulations découlant d'un jeu de données. Les outils spécifiques à la visualisation de données biologiques en forme de réseaux les plus connus sont Cytoscape ([Lopes et al., 2010](#); [Smoot et al., 2011](#)) et Pathway Studio ([Yuryev et al., 2009](#)). Cytoscape est un logiciel gratuit qui permet de visualiser sous forme de réseau de nombreux types de données (Gene Ontology, interactions, profil d'expression de gènes, liaisons binaires...) pour la biologie ([Cline et al., 2007](#)). Cytoscape est un outil communautaire qui permet l'implémentation d'applications à la source de l'outil afin d'ajouter de nouvelles fonctionnalités. Il existe plus de 230 applications gratuites réparties en une vingtaine de catégories. On y retrouve des applications créées par des bases de données d'interactions moléculaires (BioGRID, DroID, String ou PSICQUIC) pour importer directement des PPI dans Cytoscape. Il est aussi possible d'annoter un réseau de PPI à l'aide de termes GO grâce à des applications de Cytoscape (BiNGO et ClueGO ([Bindea et al., 2009](#))). Des outils de tri des réseaux comme MCODE ([Bader et Hogue, 2003](#)) recherchent les sous-réseaux de PPI impliqués dans une même voie métabolique. Il existe aussi de nombreuses applications de calcul statistique comme Cytface, une passerelle entre Cytoscape et l'outil statistique R ([Goncalves et Saez-Rodriguez, 2013](#))).

Des outils d'analyse de réseau de PPI sont disponibles pour identifier des clusters (sous réseaux) de PPI comme l'outil ClustOne ([Nepusz et al., 2012](#)), qui permet de grouper des protéines impliquées dans des fonctions biologiques communes. D'autres outils vont plus loin et permettent de comparer les clusters d'un réseau de PPI entre eux pour mettre en évidence les liens entre des fonctions biologiques (représentées par les protéines d'un cluster) : c'est le cas de l'outil Clust&See ([Spinelli et al., 2013](#)). D'autres layouts (disposition des nœuds dans le réseau) peuvent être ajoutés avec l'algorithme AllegroLayout par exemple (pour visualiser et trier plus rapidement les grands réseaux d'interaction). Cytoscape dispose aussi d'une version web implémentable ([Lopes et al., 2010](#)), par exemple PSICQUIC utilise cette version sur son interface pour visualiser les résultats d'une requête sous la forme d'un réseau. L'outil PathwayStudio, sous licence, permet aussi la création et la visualisation de réseaux biologiques. Il est possible d'importer son propre fichier d'interactions moléculaires ou de laisser l'outil rechercher des interactions grâce à sa propre base de données et à la fouille de données effectuée sur des sites de littérature comme PubMed. Récemment, PathwayStudio a proposé une version dédiée uniquement aux plantes (riz, maïs et *arabidopsis thaliana*).

Des méthodes d'analyses statistiques des composants d'un réseau d'interactions moléculaires sont utilisées afin de mettre en évidence les nœuds jouant les rôles les plus importants dans la stabilité de ce réseau. Les réseaux PPI peuvent être parcourus et triés en fonction de la notion de plus court chemin entre deux protéines grâce à l'algorithme de Dijkstra ([Dijkstra, 1971](#)). Biologiquement on recherche les nœuds (protéines ou gènes) centraux dans un réseau, pour cela on étudie notamment leur fréquence d'apparition dans les liaisons entre des groupes de molécules (Betweenness centrality) et/ou la fréquence des plus courts chemins liés à une molécule (Closeness centrality) ([Hwang et al., 2008](#); [Lim et al., 2011](#)).

Il est possible d'enrichir les résultats d'un interactome avec des protéines extérieures au jeu de données. Dans ce cas, les protéines du jeu de données sont liées avec d'autres protéines non présentes dans ce jeu de données, mais renseignées comme des interacteurs potentiels dans des bases de données spécialisées ([Lievens et al., 2010](#)). Avec cette technique, on arrive à identifier de nouvelles protéines jouant un rôle dans une fonction biologique d'intérêt, mais non présentes dans une étude. Enfin, des outils ont été développés pour aider les biologistes dans la compréhension des

réseaux. Certains outils sont dédiés à la superposition des réseaux, comme NETAL ([Neyshabur et al., 2013](#)) ou SPINAL ([Aladag et Erten, 2013](#)). Ils permettent de comparer plusieurs expérimentations et de mettre en évidence des nœuds ou des groupes de nœuds communs. La construction des réseaux d'interactions tend à ajouter de plus en plus d'informations, notamment par l'intégration de termes GO. L'outil Prodistin lie les protéines d'un réseau d'interactions en fonction de processus biologiques via des annotations avec des termes GO ceci en vue de prédire des fonctions de protéines ([Baudot et al., 2006](#); [Baudot et al., 2012](#)). Inversement, les réseaux d'interactions de protéines peuvent être utilisés pour prédire des annotations fonctionnelles, notamment par la méthode des fonctions voisines : probabilité qu'une protéine appartienne à une fonction biologique selon les fonctions dans lesquelles sont impliquées les protéines liées et voisines ([Dutkowski et al., 2013](#); [Letovsky et Kasif, 2003](#)).

Cas des ruminants

Actuellement les bases de données de PPI les plus communément utilisées rassemblent des données concernant de nombreuses espèces, dont les ruminants. Cependant, le nombre d'interactions déclarées varie grandement entre les espèces. Par exemple, avec la base IntAct, on retrouve environ 1800 PPI chez le bovin, 30 chez le mouton et 5 chez la chèvre, alors que plus de 200000 PPI sont répertoriées pour l'homme.

7 – Les outils d'analyse intégrative disponibles

Depuis plusieurs années, le nombre de données stockées dans les bases biologiques augmente de façon exponentielle ([Blake et Bult, 2006](#); [Howe et al., 2008](#)). Les résultats de séquençage, de transcriptomique et de protéomique explosent au point de faire doubler la taille d'une base de données en une année avec des données souvent dupliquées ([Byyny, 2012](#); [Schatz et Langmead, 2013](#)). La vérification des données produites et la mise à jour de l'information existante pourrait palier au problème de duplication. Cette solution est régulièrement utilisée pour mettre à jour des termes GO ([Gobeill et al., 2013](#)). On comprend ainsi que la fouille de données pour une liste d'une centaine de protéines à un temps « t » donnera souvent des résultats obsolètes à un temps « t+1 mois ». De plus, l'analyse manuelle de cette liste est inefficace compte tenu de la multiplicité des objectifs d'analyses (cf. paragraphe 5.2) et du nombre de bases de données à interroger. C'est pourquoi, les biologistes ont souvent évoqué la nécessité d'outils intégratifs tels que ceux regroupant l'interrogation de plusieurs bases de données d'interactions, la construction de réseaux et la mise en évidence des molécules centrales dans un réseau afin de visualiser les molécules essentielles aux fonctions biologiques d'un jeu de données ([Woelders et al., 2011](#)). Parmi les nombreux avantages de tels outils nous pouvons évoquer un gain de temps dans la synthèse/assemblage des connaissances disponibles, la réduction de la redondance des résultats et l'augmentation de la fiabilité des résultats en favorisant l'utilisation des informations vérifiées ou déclarées dans au moins deux sources différentes, en associant différentes méthodes d'analyses...

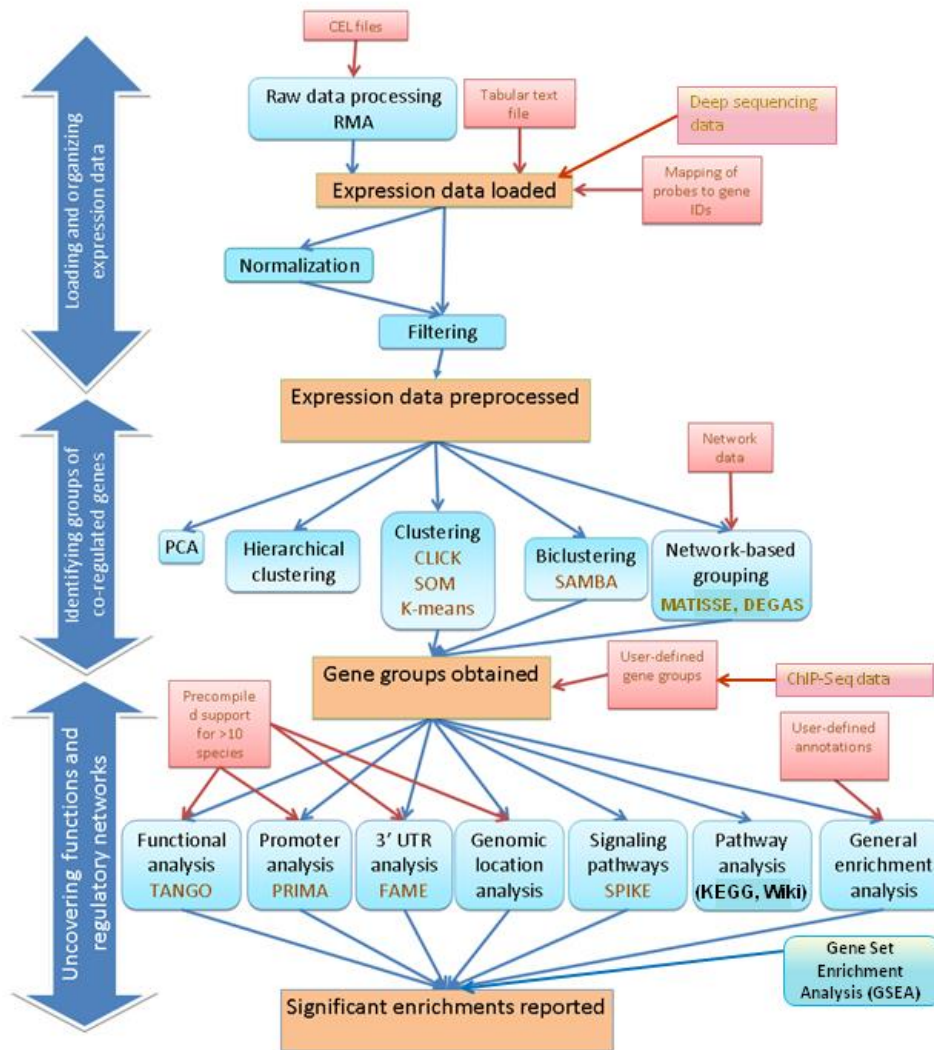


Figure 22 : Diagramme représentant le fonctionnement du workflow de l’outil Expandr (Ulitsky et al., 2010). Les rectangles rouges indiquent les fichiers d’entrée, les rectangles bleus indiquent les différents modules d’analyse et les noms des algorithmes sont écrits en orange.

Dans la littérature, le nombre d’outils tout-en-un et dédiés à l’analyse de données « omiques » augmente. Le site OmicTools référence déjà plus de 300 workflows triés dans une dizaine de catégories (génomique, protéomique, séquençage, phylogénie...). Il est impossible de tous les décrire, mais voici quelques exemples d’outils les plus couramment utilisés pour l’analyse de données « omiques ». Expandr est un outil d’analyse de données issues de puces à ADN ou de séquençage (Ulitsky et al., 2010). C’est un outil composé de plusieurs modules d’analyses (Figure 22) : l’identification de gènes différentiellement exprimés, la classification de gènes similaires dans la variation de leur abondance (clustering), l’annotation fonctionnelle par terme GO et la visualisation des gènes impliqués dans des voies métaboliques sous la forme de réseaux. Expandr est fonctionnel pour 18 espèces animales et végétales (dont l’homme, la souris, ou le riz par exemple). Babelomics est une plateforme d’analyses de données en ligne (transcriptomique et protéomique) (Medina et al., 2010). Il inclut des méthodes pour le prétraitement des données à haut débit, la normalisation, la

classification et l'expression différentielle des gènes. Il annote aussi des protéines ou des gènes, et calcule des enrichissements d'annotation. ToppGene est une plateforme d'outils en ligne pour l'analyse de grandes listes de gènes chez l'Homme ([Chen et al., 2009](#)). Il est spécialisé dans la recherche de gènes candidats pour des maladies humaines et ceci par l'annotation fonctionnelle de gènes, leurs liaisons dans des voies métaboliques et l'analyse d'interactions moléculaires. BioMyn ([Ramirez et al., 2012](#)) est un outil en ligne complet qui synthétise les informations disponibles et relatives aux gènes et aux protéines chez l'homme à partir de plus d'une douzaine de bases de données externes. Il réalise des annotations via la Gene Ontology, des classifications par famille de séquence et un affichage des domaines de protéines. Il construit aussi des réseaux de PPI pour mettre en évidence les voies métaboliques et de signalisations impliquant les gènes d'un jeu de données. Il est principalement utilisé pour des recherches d'associations de protéines impliquées dans des maladies humaines.

Outre ces outils d'analyse construits, il existe des plateformes en ligne, dédiées à la conception de workflow d'analyse de données. Parmi elles, la plateforme Galaxy ([Goecks et al., 2010](#)) est accessible à des utilisateurs non programmeurs. Elle propose un large panel d'outils paramétrables facilement, qui s'assemblent et s'utilisent uniquement via des clics de souris. Galaxy permet de partager des workflows préconstruits par la communauté scientifique pour avoir un protocole d'analyse de données toujours identique. Les composants de Galaxy sont téléchargeables localement pour autoriser la création de serveurs Galaxy privés. Malgré sa simplicité, il est recommandé d'effectuer une formation pour démarrer des projets avec Galaxy. Le système d'exploitation gratuit, BioLinux ([Field et al., 2006](#)), est entièrement dédié à l'analyse de données biologiques. Il met à disposition 260 outils préinstallés comme le Blast en local, Cytoscape, le logiciel de calcul statistique R (avec de nombreux packages ; <https://www.r-project.org/>), Bioperl ([Stajich et al., 2002](#)) ou encore Primer3 ([Untergasser et al., 2012](#)). Cependant, BioLinux reste un système d'exploitation réservé à un public de programmeurs car son interface et l'utilisation de certains outils nécessitent un paramétrage en ligne de commandes.

Bien que de nombreux outils soient disponibles dans chaque domaine de la bioinformatique présenté, aucun d'eux ne s'appliquent à l'étude des données expérimentales de ruminants. Il serait possible d'étudier chaque domaine indépendamment en suivant la stratégie d'analyse de données de protéomiques et en utilisant les orthologie et homologie entre espèce. Cependant, le temps consacré à une telle étude serait plus important que celui consacré à la conception d'un nouvel outil unique et dédié aux ruminants.

8 – Conclusion et objectifs du travail de thèse

Il ressort de notre étude bibliographique que les connaissances associées à la croissance précoce des tissus adipeux et musculaires restent fragmentaires. En particulier les mécanismes qui contrôlent la détermination et la prolifération des progéniteurs musculaires et adipeux. De même, les protéines qui participent aux dialogues « muscle-TA » restent largement à décrire. Nous avons fait l'hypothèse que l'intégration de données de génomique disponibles chez le fœtus bovins et l'acquisition de nouvelles données chez des fœtus avant l'âge de 180 jpc (au-delà le nombre de fibres musculaires est fixé), permettraient de mieux comprendre les mécanismes mis en jeu lors de la

croissance *in vivo* des tissus à la fois par hyperplasie et par hypertrophie. Cependant, une fouille de données de génomique efficace chez le ruminant requiert des outils informatiques dédiés qui restent à développer. Dans ce contexte l'objectif de la thèse **était de concevoir un serveur web pour la fouille de données de génomique chez le bovin, afin d'améliorer les connaissances sur les mécanismes moléculaires qui participent à la croissance (par hyperplasie et hypertrophie) des tissus musculaires et adipeux chez le fœtus bovin.**

Pour ce faire, notre étude a été découpée en deux parties. Dans un premier temps, nous avons développé un service web de fouille de données de génomique dédié aux ruminants, dont le bovin. La pertinence des informations et des résultats produits par l'outil a été testée et comparée à ceux produits par d'autres ressources bioinformatiques. Dans un second temps, nous avons caractérisé les protéomes fœtaux des tissus musculaires et adipeux bovins. Pour cela, nous avons utilisé des données de protéomique disponibles au laboratoire, que nous avons complétées par une étude du protéome de deux TA et de deux muscles prélevés à 140 jpc.

MATERIELS ET METHODES

Matériels et méthodes

1 – Données expérimentales de protéomique

Des données de protéomique ont été utilisées pour la validation de l'outil ProteINSIDE et son application pour enrichir les connaissances sur la croissance fœtale des tissus musculaire et adipeux chez le bovin (cf. Objectifs et Résultats). Ces données sont issues de deux expérimentations. La première concerne la cinétique de mise en place des tissus musculaire (muscle *Semitendinosus*, ST correspondant au rond de gîte en boucherie) et adipeux (TA : périrénal) à 4 ou à 5 âges fœtaux chez le bovin. Les protéomes de ces deux tissus ont été analysés respectivement dans les thèses de Thibault Chaze ([Chaze, 2007](#)) et Hager Taga ([Taga, 2011](#)) dont les résultats ont été présentés dans la partie étude bibliographique. Une seconde expérimentation porte sur la description des protéomes musculaires et adipeux à 140 jours de vie fœtale ([Gonzalez et al., 2013](#)). Dans cette partie nous présenterons les protocoles expérimentaux utilisés pour obtenir et analyser ces données.

1.1 – Animaux et échantillons

Cinétique d'ontogenèse des tissus

Des vaches primipares de races Charolaise et Blonde d'Aquitaine ont été inséminées artificiellement (à l'unité expérimentale de l'INRA UMRH1213). Quarante et un fœtus (20 mâles et 21 femelles) ont été obtenus aux stades 60, 110, 180, 210 et 260 jours post-conception (jpc ; [Figure 23](#)). Ces stades ont été choisis en fonction des connaissances acquises sur la myogenèse et l'adipogenèse chez le bovin ([Bonnet et al., 2010](#); [Picard et al., 2003](#)). Le stade 60 jpc correspond à la prolifération des myoblastes primaires et au début de prolifération des myoblastes secondaires ; le stade 110 jpc est caractérisé par la différenciation des myotubes primaires et la prolifération des myoblastes de seconde et troisième générations ; 180 jpc est un stade clé marqué par la fin de la prolifération des myoblastes ; les stades 210 et 260 jpc correspondent à la différenciation et à la maturation de l'ensemble des fibres musculaires. Dans cette expérimentation, le régime alimentaire des vaches a été le même durant toute la période de gestation. Les échantillons fœtaux de TA périrénal et de muscle ST ont été prélevés juste après abattage des mères.

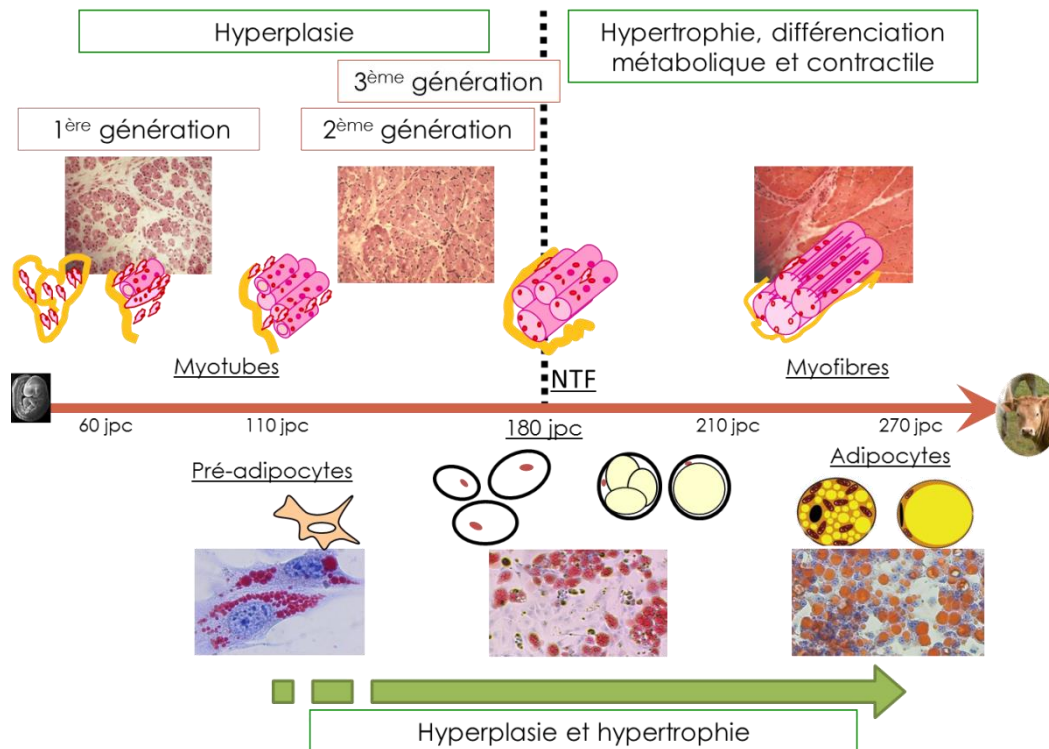


Figure 23 : Chronologie de la croissance des tissus musculaire et adipeux au cours de la vie fœtale chez le bovin. Les cellules musculaires prolifèrent jusqu'à fixation du nombre total de fibres musculaires (NTF), se différencient et acquièrent des fonctions biologiques. Les cellules adipeuses prolifèrent et se différencient continuellement, bien que la prolifération ait lieu essentiellement en période périnatale (pour revue ([Bonnet et al., 2010](#); [Picard et al., 2002](#))).

Description des protéomes musculaires et adipeux à 140 jours de vie fœtale

Dans le projet « Fœtal Programming », les vaches en gestation (croisement de race British White et British Freisian) ont reçu différents régimes alimentaires alternant nutrition normale (couvrant 100% des besoins en énergie et protéines ; C) et sous-nutrition/restriction (seulement 60% des besoins en énergie et protéines étaient couverts ; R) selon le protocole décrit en [Figure 24](#). Trois lots de 5 ou 6 vaches ont été conduits : un lot contrôle qui a reçu l'alimentation C durant toute la gestation, un lot CR qui a reçu l'alimentation C de 0 à 30 jpc puis l'alimentation R de 30 à 85 jpc puis de l'alimentation C de 85 à 140 jpc, et le lot RR qui a reçu l'alimentation R de 30 à 140 jpc. Dans notre étude nous avons considéré uniquement l'âge fœtal de 140 jpc. Ce stade est caractérisé par la prolifération et la différenciation des deux tissus ([Bonnet et al., 2010](#)). Après abattage des mères, les prélèvements ont été effectués sur les TA omental et périrénal et sur les muscles *semitendinosus* (ST) et le *longissimus thoracis* (LT). Pour chaque animal, le poids du fœtus, de certains organes (cœur, reins, intestins...), des deux muscles et des quatre tissus adipeux ont été référencés. Pour les deux expérimentations, les tissus prélevés ont été découpés et congelés dans l'azote liquide, puis stockés à -80°C pour les mesures d'abondance des ARNm et des protéines.

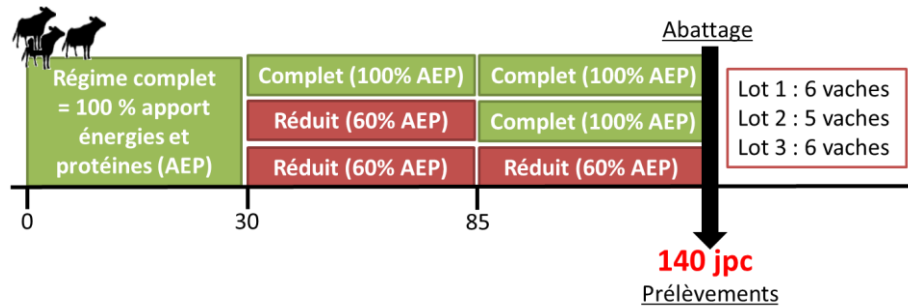


Figure 24 : Protocole nutritionnel de la seconde expérimentation (Gonzalez *et al.*, 2013).

1.2 – Protéomique

Afin d'identifier les protéines impliquées dans la prolifération et/ou dans la différenciation cellulaire au cours de la croissance fœtale, les profils protéiques des échantillons des deux tissus ont été analysés par protéomique. Le protocole suivi consiste en la séparation des protéines par électrophorèse bidimensionnelle, leur analyse différentielle par analyse d'image et l'identification des protéines d'intérêt par spectrométrie de masse.

Extraction et identification des protéines de l'ontogenèse des tissus

Les protéines des échantillons issus des deux expérimentations ont été extraites selon le protocole décrit par Bouley *et al.* (Bouley *et al.*, 2004) pour le muscle et Taga *et al.* (Taga *et al.*, 2012) pour les TA. Elles ont ensuite été séparées par électrophorèse bidimensionnelle (2D) selon la méthode décrite par Bouley *et al.* (Bouley *et al.*, 2004). La méthode consiste à séparer les protéines d'un échantillon en deux dimensions (Figure 25) : en fonction de leur point isoélectrique pI (isoélectrofocalisation, IEF) puis selon leur poids moléculaire (par SDS-PAGE). Pour l'IEF, les protéines ont été séparées selon un gradient de pH de 4-7 pour la première expérimentation et de 5-8 pour la seconde expérimentation. Un gel 2D a été réalisé pour chaque échantillon. Les gels obtenus ont été scannés puis analysés informatiquement (Figure 26) afin de détecter et de quantifier les spots protéiques à l'aide du logiciel SameSpots ProGenesis (v. 3 puis v. 5 pour la seconde expérimentation; Totallab Software). Dans la première expérimentation, pour la mise en évidence des spots protéiques différant entre deux situations, les gels 2D de chaque animal ont été comparés à un gel de référence réalisé à partir d'un mélange des protéines des âges fœtaux considérés pour chacun des deux tissus. Ce gel de référence est numérisé dans les mêmes conditions que les autres gels et tous les gels à analyser sont alignés sur ce gel de référence par le logiciel SameSpots. L'abondance des spots est quantifiée et exprimée en % de volume puis normalisée par rapport à la moyenne des abondances de tous les spots.

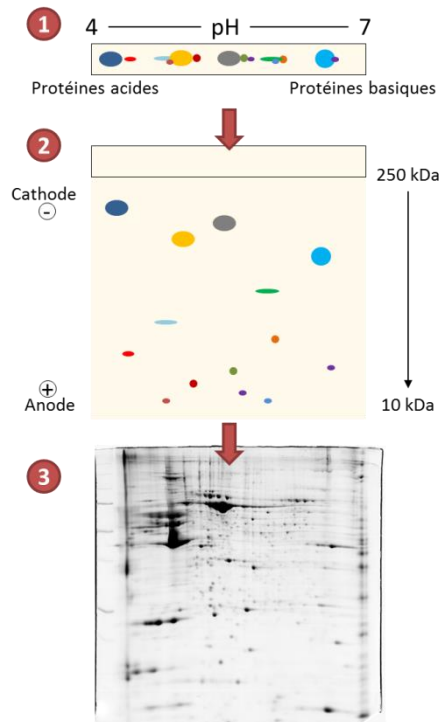


Figure 25 : Schéma explicatif d'une électrophorèse bidimensionnelle. Dans un premier temps (1) les protéines migrent dans un gradient de pH et sont séparées en fonction de leur point isoélectrique (pH auquel la charge de la protéine est nulle) puis (2) elles sont séparées en fonction de leur poids moléculaire dans un gel SDS-PAGE. Après coloration au bleu de coomassie colloïdal G250 (3), les gels sont scannés et les spots protéiques sont analysés par analyse d'image.

L'identification des spots est effectuée par spectrométrie de masse sur la plateforme « Exploration du Métabolisme : des gènes aux nutriments » du centre INRA de Theix. Les spots découpés à partir des gels sont traités (décoloration, déshydratation puis hydrolyse en peptides par la trypsine) puis analysés par MALDI-TOF (Matrix Assisted Laser Desorption/Ionization – Time Of Flight), dans la première expérimentation, ou par LC-MS (Liquid chromatography - Mass Spectrometry : chromatographie liquide et spectrométrie de masse), dans la deuxième expérimentation (Figure 26).

Ces méthodes de spectrométrie de masse sont basées sur la mesure de la masse du mélange peptidique, obtenu par la digestion enzymatique de la protéine inconnue, selon son temps de vol. On compare les empreintes peptidiques massiques obtenues à celles stockées dans la base de données RefSeq *Bos taurus* du NCBI en utilisant le logiciel MASCOT (v. 2.2 ; Matrix Science software ; (Koenig *et al.*, 2008) ; http://www.ncbi.nlm.nih.gov/genome/annotation_euk/Bos_taurus/103/).

Lors de la première expérimentation avec une analyse MALDI-TOF, seules les protéines ayant le meilleur score d'identification ont été sélectionnées. Pour la seconde expérimentation avec une analyse par LC-MS qui révèle plus de protéines par spot, j'ai réalisé un script pour trier les protéines identifiées sous chaque spot en fonction du nombre de peptides et du score de l'identification. Contrairement aux données précédentes, toutes les protéines identifiées avec au moins 3 peptides ont été sélectionnées pour les protéomes.

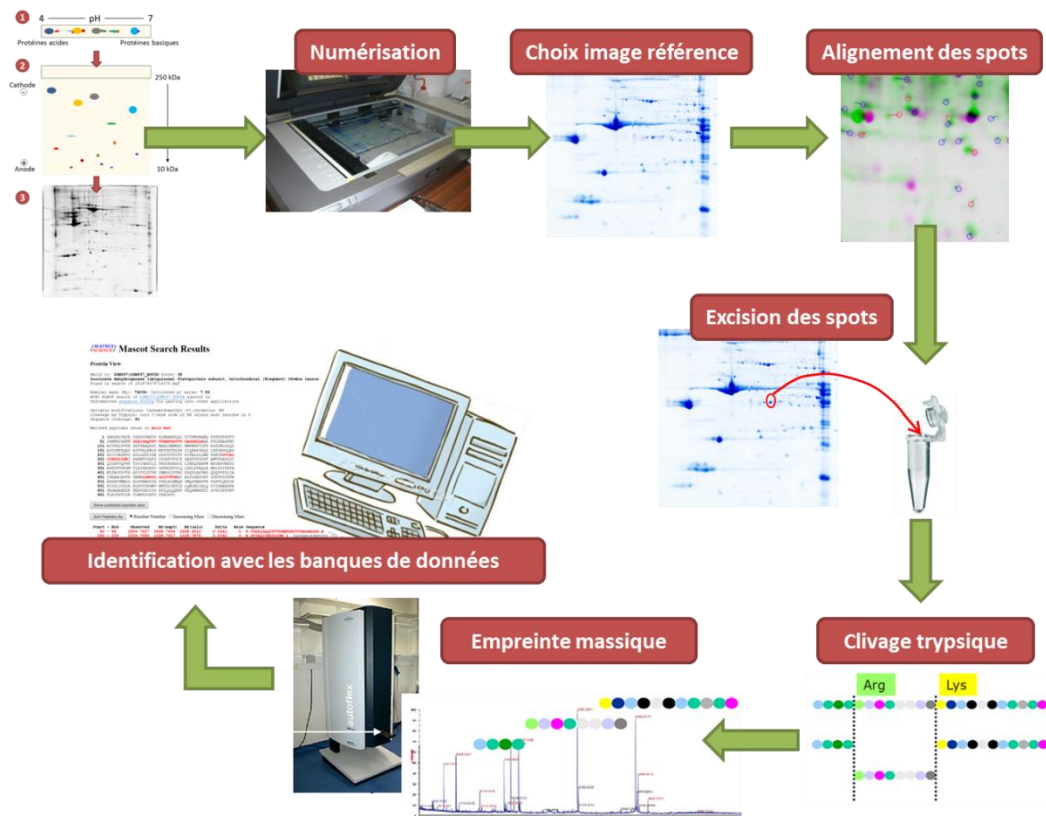


Figure 26 : Différentes étapes du protocole d'identification des protéines suivi au laboratoire Amuvi. Après la séparation des protéines par gel 2D, celui-ci est scanné pour être analysé informatiquement afin de révéler les spots protéiques communs entre plusieurs gels. Les spots d'intérêt sont extraits et les protéines sont digérées par action de la trypsine. Les échantillons sont ensuite analysés par spectrométrie de masse et les résultats sont comparés à une base de données pour identifier les protéines.

1.3 – Analyses bioinformatiques et clustérisations

Données de l'ontogénèse des tissus (première expérimentation)

A l'issue de l'identification par spectrométrie de masse, deux jeux de données de 143 et 120 protéines respectivement pour le TA et le muscle, ont été obtenus ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)). À titre d'exemple, une analyse en composantes principales (ACP) a été réalisée pour les données du muscle. Les abondances des protéines discriminent les âges étudiés ([Figure 27 – A](#)). Aussi, afin d'identifier les protéines qui ont des profils d'abondance similaires en fonction des âges fœtaux, les protéines ont été clustérisées en fonction de leurs abondances ([Meunier et al., 2007](#)) aux différents âges fœtaux ([Figure 27 – B](#)). Cette clustérisation des abondances des protéines a été réalisée pour chaque tissu avec le logiciel PermutMatrix ([Caraux et Pinloche, 2005](#)). Les paramètres de classification considérés étaient le coefficient de Pearson et l'index de Jacard (respectivement pour les données quantitatives et qualitatives) pour la mesure de distance, et la méthode de Ward pour la mesure d'agrégation et l'obtention du dendrogramme. Les clusters de protéines ont ensuite été analysés à l'aide d'outils bioinformatiques pour mettre en évidence des fonctions biologiques d'intérêt pour la croissance fœtale des tissus ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)). Pour chaque protéine, la stratégie bioinformatique utilisée en 2007 pour le muscle et 2010 pour le TA, était de rechercher la fonction biologique sur la base de données UniProt

(UniProt, 2015), d'annoter la protéine avec des termes GO grâce à DAVID pour détecter les processus biologiques (Huang *et al.*, 2009) et de déterminer les voies métaboliques empruntées par la protéine avec KEGG (Kanehisa *et al.*, 2015). Pour mes travaux de thèse j'ai utilisé les identifiants des 143 et 120 protéines adipeuses et musculaires

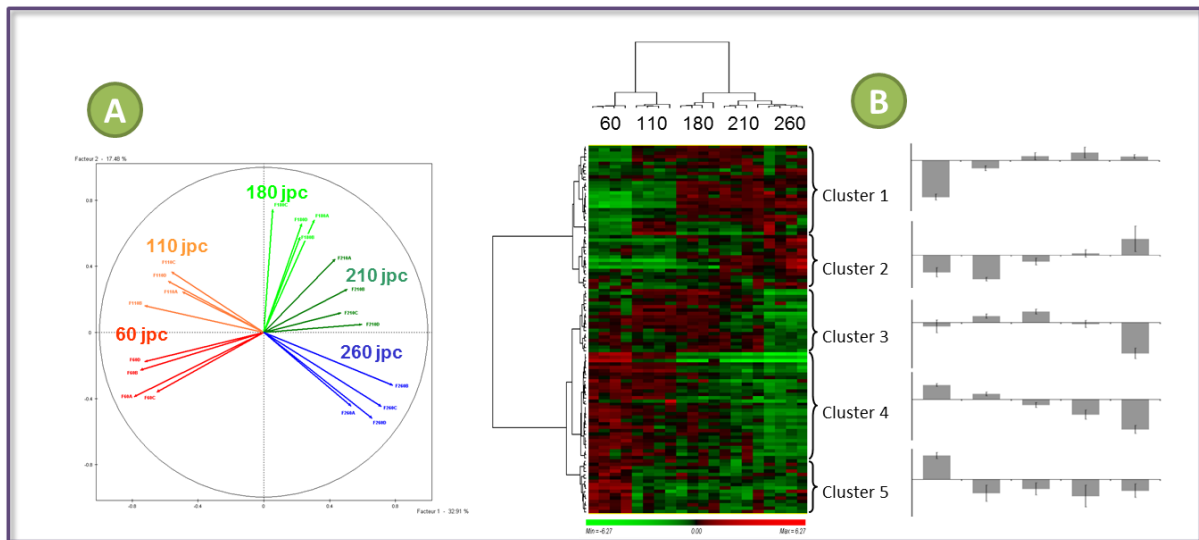


Figure 27 : Analyse des données de protéomique de la myogenèse du muscle *Semitendinosus* de fœtus bovins de race charolaise (Chaze *et al.*, 2008). (A) Analyse en composante principale illustrant la distinction des 5 stades étudiés (n=5 fœtus/stade) à partir de l'abondance des protéines révélées. (B) Analyse hiérarchique illustrant l'abondance de 120 protéines communes aux 5 stades étudiés et permettant la distinction de 5 profils d'évolution d'abondance de protéines différents (5 clusters). En vert les protéines peu abondantes, en rouge les protéines très abondantes.

Données des protéomes fœtaux à 140 jpc (seconde expérimentation)

A l'issue de l'identification par spectrométrie de masse, deux jeux de données de 752 et 515 protéines bovines ont été obtenus respectivement pour le TA et le muscle. Parmi ces protéines adipeuses et musculaires, 705 et 490 ont des orthologues chez l'homme. Une analyse de variances (Anova) sous R (version 3.2) a été réalisée afin d'identifier les spots dont l'abondance était différentielle en fonction de la nutrition maternelle, du site anatomique des TA (périrénal *versus* omental) ou des muscles (LT *versus* ST) ou de l'interaction entre ces deux facteurs. La *p*-value inférieure a été corrigée par le test de Benjamini et Hochberg a (Benjamini et Hochberg, 1995) et les abondances ont été considérées différentielles pour les *p*-value < 0.10.

2 – Conception de l'outil bioinformatique ProteINSIDE

ProteINSIDE a été divisé en trois parties (Figure 28) : le workflow, la base de données et l'interface web en ligne. L'interface web est le support de ProteINSIDE et permet aux utilisateurs de consulter le site (création d'analyses, accès aux résultats, mises à jour...). La base de données est le support des informations de ProteINSIDE (analyses, résultats, données biologiques...). Le workflow

est le support des modules d'analyses de ProteINSIDE. Il analyse les identifiants issus de 6 espèces : la souris, le rat, l'homme, le bovin, la chèvre et le mouton.

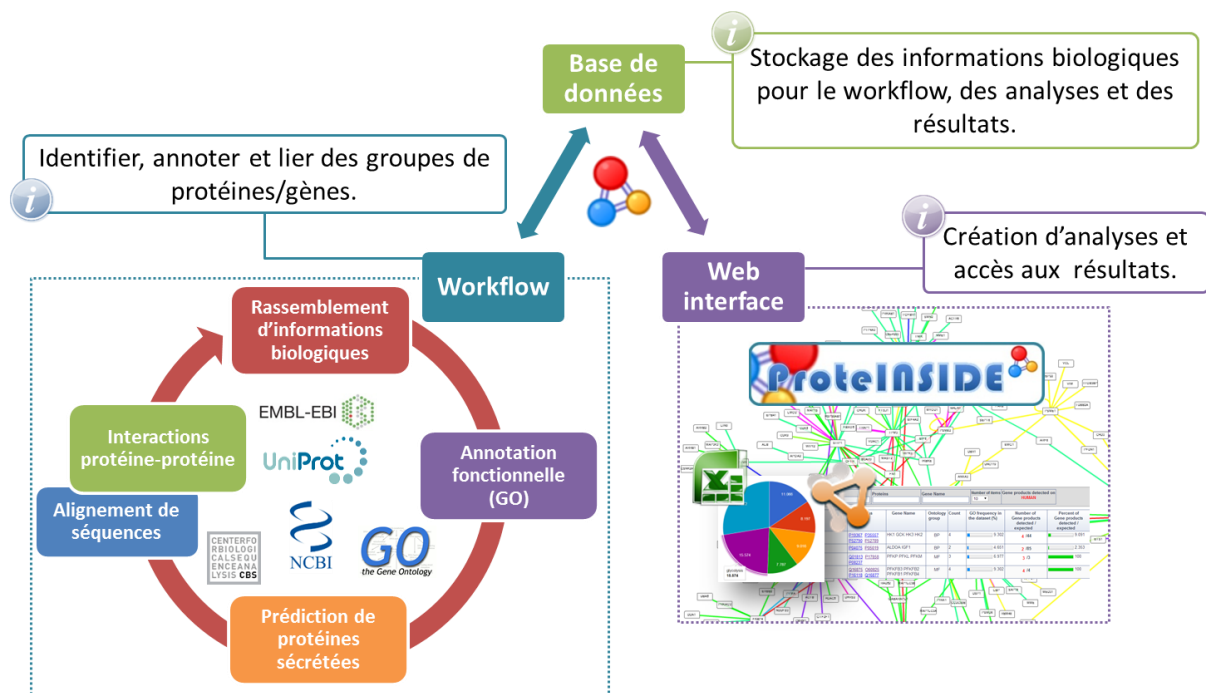


Figure 28 : Organigramme des trois parties de ProteINSIDE : la base de données, l'interface web et le workflow. Ce dernier est divisé en quatre modules d'analyse.

2.1 – Développement des modules d'analyse

Le workflow de ProteINSIDE est divisé en quatre modules d'analyse (Figure 28) :

- « ID mapping », il rassemble les connaissances biologiques disponibles pour des protéines ou des gènes
- « Gene Ontology (GO) », il utilise des termes GO pour annoter les fonctions des protéines
- « Secreted Proteins », il prédit les protéines sécrétées
- « Protein Interaction », il recherche les interaction entre protéines

Tous les modules de ProteINSIDE ont été programmés en langage Perl (version 5.10.1 et des packages CPAN (Comprehensive Perl Archive Network)) et en R (version 3.2.0). Le package Perl DBI (<http://search.cpan.org/~timb/DBI-1.634/DBI.pm>) a été utilisé dans tous les programmes pour interagir avec la base de données de ProteINSIDE. Le package Perl Excel-Writer (<http://search.cpan.org/~jmcnamara/Excel-Writer-XLSX-0.86/lib/Excel/Writer/XLSX.pm>) a été employé pour créer les fichiers de résultats au format Excel dans tous les modules de ProteINSIDE.

Module « ID Mapping »

Ce module est indispensable à ProteINSIDE. Il reconnaît les protéines et/ou les gènes utilisés dans une analyse à partir des identifiants (ID) envoyés. La base de données de ProteINSIDE possède une table qui référence tous les ID de protéines et de gènes pour les 6 espèces de l'outil. Cette table rassemble pour chaque ID les connaissances biologiques qui lui sont associées : fonction biologique, localisation chromosomique du gène, orthologie et homologie entre espèces, spécificité tissulaire, localisation cellulaire, d'autres ID connus dans les bases de données biologiques et la séquence. De cette façon, quelle que soit l'analyse créée, ProteINSIDE retrouve les ID utilisés même s'ils proviennent de plusieurs bases différentes. Par exemple, pour l'adiponectine plusieurs ID sont utilisables avec ProteINSIDE : Q15848 (UniProt), ADIPO_HUMAN (UniProt / NCBI Protein), ADIPOQ (Gene Name), gi|62022275 (Gene Identifier) ou ENSG00000181092 et ENST00000444204 (respectivement Ensembl gène et transcrit). Si l'ID n'est pas retrouvé, il est toujours possible d'utiliser un convertisseur intégré à ProteINSIDE, nommé ProteCONVERT. Il interroge directement le service « ID Mapping » de la base UniProt. Ce service recherche, pour un ID, tous les ID connus de toutes les bases de données biologiques. Les séquences des protéines sont récupérées *via* la fiche de la protéine sur la base de données UniProt ([UniProt, 2015](#)). ProteINSIDE crée des liens vers les fiches de gènes (NCBI ; ([Coordinators, 2015](#))) et de protéines (UniProt), et rassemble tous les résultats dans un fichier Excel et les séquences protéiques au format FASTA dans un fichier téléchargeable.

Module « GO »

L'annotation par des termes GO est effectuée par l'interrogation de chaque ID sur l'outil QuickGO ([Binns et al., 2009](#)) et selon les connaissances disponibles pour l'espèce en cours d'analyse et répertoriées dans AmiGO ([Carbon et al., 2009](#)). Les termes GO et les protéines associées aux termes sont référencés. L'enrichissement de chaque terme GO est calculé en fonction des protéines qu'il annote dans le jeu de données et en fonction de toutes les protéines qu'il annote dans cette espèce. L'enrichissement des termes est calculé avec un script R selon un test exact de Fischer par tableau de contingence (cf. Résultats paragraphe 1.3). Le résultat de ce test identifie les termes GO les plus significatifs pour le jeu de données (le score/*p*-value résultant du test est aussi corrigé par la méthode Benjamini & Hochberg ([Benjamini et Hochberg, 1995](#))). Des graphiques des proportions de protéines annotées, pour chaque groupe fonctionnel (Biological Process, Molecular Function et Cellular Component) et termes GO, sont créés. Nous avons aussi implémenté une visualisation des termes GO sous la forme de réseaux en arbres de phylogénie grâce à l'interface de Cystocape web ([Lopes et al., 2010](#)). Dans ce réseau, seuls les termes GO obtenus lors de l'annotation sont présents et des nuances de couleurs mettent en évidence les termes annotant de beaucoup à peu de protéines du jeu de données. Les scores d'enrichissement des termes sont indiqués. Un tableau de résultats au format Excel est créé et rassemble tous les termes GO, protéines annotées et résultats statistiques.

Module « *Secreted Proteins* »

La prédiction des protéines potentiellement sécrétées repose sur l'analyse de leur séquence protéique avec le logiciel SignalP ([Petersen et al., 2011](#)). Afin d'appuyer cette prédiction, ProteINSIDE affiche les termes GO liés à la fonction de sécrétion pour chaque protéine prédite comme ayant un peptide signal. La base de données de ProteINSIDE possède un dictionnaire d'environ 1000 termes GO liés aux localisations cellulaires des protéines, triés en fonction de toute l'arborescence de la GO actuelle. En plus des termes GO, le module répertorie la localisation cellulaire des protéines selon les annotations UniProt précédemment référencées par le module « ID Mapping ». Enfin une prédiction de cette localisation cellulaire est réalisée avec le logiciel TargetP ([Emanuelsson et al., 2000](#)). Des scores calculés par les deux outils SignalP et TargetP ont aussi été ajoutés pour informer les utilisateurs sur la qualité de la prédiction. Tous les résultats sont téléchargeables sous la forme d'un fichier Excel.

Module « *Protein Interaction* »

Les interactions entre protéines (PPI) sont obtenues grâce à des requêtes avec le service web Psiquic ([Aranda et al., 2011](#)). Ce service peut actuellement interroger 32 bases de données de PPI. Pour chaque protéine du jeu de données, le module recherche, sur chaque base de données de PPI sélectionnée, les interactions disponibles pour la protéine. Ces interactions sont sélectionnées en fonction de leur méthode de détection. ProteINSIDE garde uniquement les méthodes de détection expérimentales. Dans le cas des PPI entre protéines d'un jeu de données, ProteINSIDE vérifie si l'autre protéine de l'interaction est bien présente dans le jeu de données. Dans le cas des PPI étendues, c'est-à-dire de la recherche d'interactions entre les protéines du jeu de données avec des protéines extérieures au jeu de données, ProteINSIDE vérifie si la protéine extérieure au jeu de données est connue dans l'espèce d'étude (ou de l'espèce choisie pour rechercher les PPI). L'utilisateur peut choisir de rechercher des PPI étendues dans une espèce différente choisie par ProteINSIDE. Dans ce cas, ProteINSIDE effectue un Blastp sur toutes les protéines du jeu de données pour déterminer l'espèce la plus similaire à celle de l'étude en cours. Ensuite, les PPI étendues sont sélectionnées sur cette espèce. Le Blastp est effectué avec une version locale (version 2.2.29 du NCBI) contre la base de données UniProt/SwissProt (version 2.2.21). Si une même PPI est retrouvée plusieurs fois avec des méthodes de détections différentes, elles sont toutes sélectionnées dans les résultats. S'il y a des doublons, ils sont comptabilisés et appuient la robustesse d'une PPI. Un script R calcule les tests de centralités (« betweenness » (intermédiaire) et « closeness centrality » (proximité)) pour toutes les protéines impliquées dans les PPI en utilisant le package « tnet » ([Tore, 2009](#)). Les résultats d'interactions sont disponibles sous la forme de tableaux ou de réseaux de protéines (grâce à Cytoscape web). Les réseaux peuvent être triés et visualisés sous différentes formes afin d'affiner l'affichage des interactions.

Les résultats produits par chaque module sont téléchargeables : les tableaux au format Excel, les images aux formats jpg ou pdf, les réseaux sous différents formats (graphml, xgmml, sif...) réutilisables avec d'autres logiciels de visualisation comme Cytoscape bureau ([Smoot et al., 2011](#)).

2.2 – Développement web de ProteINSIDE

La base de données de ProteINSIDE, codée en MySQL, est composée de 21 tables pour stocker les informations de l'outil comme : les analyses, les résultats, les utilisateurs ou les articles tutoriaux du site. Elle stocke aussi une version triée des connaissances biologiques disponibles pour chaque espèce. Elle permet une reconnaissance rapide des ID envoyés dans l'outil et une réduction de la charge du serveur lors de l'ouverture en ligne des résultats. La base de données stocke les paramètres des analyses effectuées avec l'outil et les résultats de ces analyses. Toutes les données concernant les analyses et les résultats sont cryptées.

L'interface web en ligne est programmée en PHP, HTML et JavaScript. Elle est nécessaire à la création d'une nouvelle analyse, à la visualisation des résultats et à l'information des utilisateurs sur les mises à jour de l'outil. Tous les tableaux et graphiques du site sont réalisés avec un module Google nommé Google Charts (<https://developers.google.com/chart/>). Ce module Google permet une visualisation dynamique des résultats pour un tri instantané des informations. L'interface dispose aussi de tutoriaux pour comprendre le fonctionnement des analyses et l'interprétation des résultats.

ProteINSIDE a dans un premier temps été hébergé sur un serveur avec une architecture Debian (version 6 Squeeze), un processeur Intel Xeon E5603 (4 cœurs) et 4 Go de mémoire vive. ProteINSIDE a ensuite migré vers un serveur plus performant (64 cœurs et 128 Go de mémoire vive).

RESULTATS

Résultats

1 – Développement d'un outil de fouille de données de génomique dédié aux ruminants : ProteINSIDE

1.1 – Introduction

L'objectif des scientifiques travaillant sur l'efficacité productive des ruminants (bovins, ovins ou caprins) et la qualité de leurs produits (viande, lait ...) est de comprendre quels gènes ou quelles protéines contrôlent le métabolisme des nutriments et leur utilisation par les tissus, ou quels gènes et quelles protéines contrôlent la croissance des tissus ([Bonnet et al., 2010](#)). Pour y parvenir, différentes méthodes sont mises en œuvre comme l'annotation du génome, le séquençage à haut débit et la quantification de l'expression des gènes. La masse de données produites par ces études de génomique augmente continuellement ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Hudson et al., 2013](#); [Lehnert et al., 2007](#); [Picard et al., 2011](#); [Sudre et al., 2003](#); [Taga et al., 2012](#)). Il est donc nécessaire d'analyser ces données pour générer des informations biologiques et des connaissances ([Woelders et al., 2011](#)). C'est possible grâce à des outils informatiques, utilisant des identifiants (ID) souvent différents pour une même protéine ou un gène. Ces ID sont aussi utiles pour interroger des bases de données comme UniProtKB ([UniProt, 2015](#)) ou NCBI ([Coordinators, 2015](#)), qui stockent une grande quantité d'informations à lire et à trier. De plus, la majorité des outils bioinformatiques sont spécifiques à un type d'analyse : comme l'annotation selon la Gene Ontology (GO ; [Ashburner et al., 2000](#)), la prédiction de peptides signaux ([Klee et Ellis, 2005](#)) ou l'identification d'interactions moléculaires ([Braun et Gingras, 2012](#)). Des workflows intégrant plusieurs types d'analyses sont disponibles (cf. Étude bibliographique, paragraphe 7), mais sont restreints à une espèce (l'homme, la souris, la drosophile, *Arabidopsis thaliana*, *Escherichia coli*...) et sont donc peu, ou pas, appropriés pour l'analyse des données de génomique provenant des ruminants. À ce jour, il n'y a donc pas de workflow pour une analyse intégrative et efficace des données de génomique provenant des ruminants. De plus, et contrairement à l'homme ou aux espèces modèles comme la souris ou le rat, les ruminants sont moins bien annotés (il s'agit souvent d'annotations électroniques) et les informations stockées dans les bases de données sont peu revues par des experts. Souvent les scientifiques travaillant chez les espèces peu annotées, mais phylogénétiquement proches, utilisent les orthologues ou des homologues dans le but d'utiliser les connaissances biologiques disponibles dans les espèces mieux annotées. Des liens d'orthologie et d'homologie existent entre espèces et sont prouvés. De tels liens sont référencés dans les bases OrthoDB ([Kriventseva et al., 2015](#)), dédiée aux orthologies, et HomoloGene ([Coordinators, 2015](#)), dédiée aux homologies entre espèces. L'intégration, dans un workflow, des liens d'orthologie et d'homologie entre les protéines/gènes de ruminants et les espèces humaines et modèles, n'a jamais été faite.

Dans ce contexte, nous avons décidé de concevoir un workflow pour la fouille de données de génomique dédiée aux ruminants : ProteINSIDE. **L'objectif était de proposer un outil avec lequel la synthèse des connaissances biologiques ou la prédiction de connaissances sont effectuées en une seule requête pour une liste de gènes ou de protéines bovine, ovine ou caprine.** ProteINSIDE a été réfléchi pour compléter les informations manquantes chez les ruminants en utilisant, par homologie, les annotations disponibles chez les espèces monogastriques mieux annotées (souris, rat et homme).

ProteINSIDE effectue ce lien d'homologie grâce à la base HomoloGene du NCBI. Ainsi, un utilisateur peut créer une analyse chez l'homme en utilisant des ID bovins. Dans ce cas, les ID bovins détectés comme homologues chez l'homme sont gardés (les autres sont écartés de l'analyse). Cette détection d'homologie se fait par comparaison dans la base de données de ProteINSIDE qui possède une version locale des homologues référencés dans la base HomoloGene. Les résultats produits rendent compte de l'état des connaissances chez l'homme. La version bêta de ProteINSIDE a été présentée à la deuxième conférence internationale de bioinformatique et d'ingénierie biomédicale (International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO ; ([Kaspric et al., 2014](#))), et la version aboutie du service web a été publiée dans le journal Genomics and Computational Biology (GCB ; ([Kaspric et al., 2015](#))).

1.2 – La version bêta de ProteINSIDE

La première version de ProteINSIDE a été mise en ligne en octobre 2013 et rendue immédiatement accessible gratuitement. Cette version comportait les quatre principaux modules de fouilles de données et de prédictions de ProteINSIDE ([Figure 29](#)) : la recherche des connaissances biologiques pour les ID d'un jeu de données (module « ID mapping »), l'annotation fonctionnelle par des termes GO (module « Gene Ontology »), la prédiction de protéines sécrétées par un peptide signal (module « Secreted Proteins ») et la recherche d'interactions entre protéines du jeu de données (module « Protein Interaction »). ProteINSIDE recherchait les interactions non seulement entre les protéines d'un jeu de données (PPI interne), mais aussi entre les protéines du jeu de données et des protéines qui n'étaient pas dans le jeu de données (PPI étendue). Pour tester l'outil, nous avons constitué un jeu de données de 133 ID UniProt de protéines (dont un doublon). Nous avons sélectionné une majorité de protéines impliquées dans la glycolyse, le cycle de Krebs et des protéines connues pour être des hormones ou être sécrétées (comme l'insuline, l'hormone Natriuretic peptides A et la Resistine). Ce jeu de données, dont les fonctions biologiques des protéines sont connues dans la littérature (par exemple l'adiponectine, l'énolase 3), a permis de tester l'importation des informations biologiques, l'annotation des protéines par les termes GO et la prédiction des protéines sécrétées. Les interactions entre protéines ont aussi été testées, car de nombreuses protéines sont impliquées dans les mêmes fonctions biologiques. Les tests ont été réalisés avec les deux méthodes proposées par ProteINSIDE : une analyse basique (paramètres automatiques) et une analyse personnalisée (l'utilisateur choisit les paramètres d'analyses). Les résultats obtenus pour ce fichier « test » ont été mis en ligne afin de montrer les fonctionnalités de l'outil. ProteINSIDE a directement été conçu pour analyser des ID de 6 espèces : humain, murin, rat, bovin, caprin et ovin.

La version bêta de l'outil et ses performances ont fait l'objet d'une communication écrite longue et d'une communication orale à la 2^{ème} conférence IWBBIO.

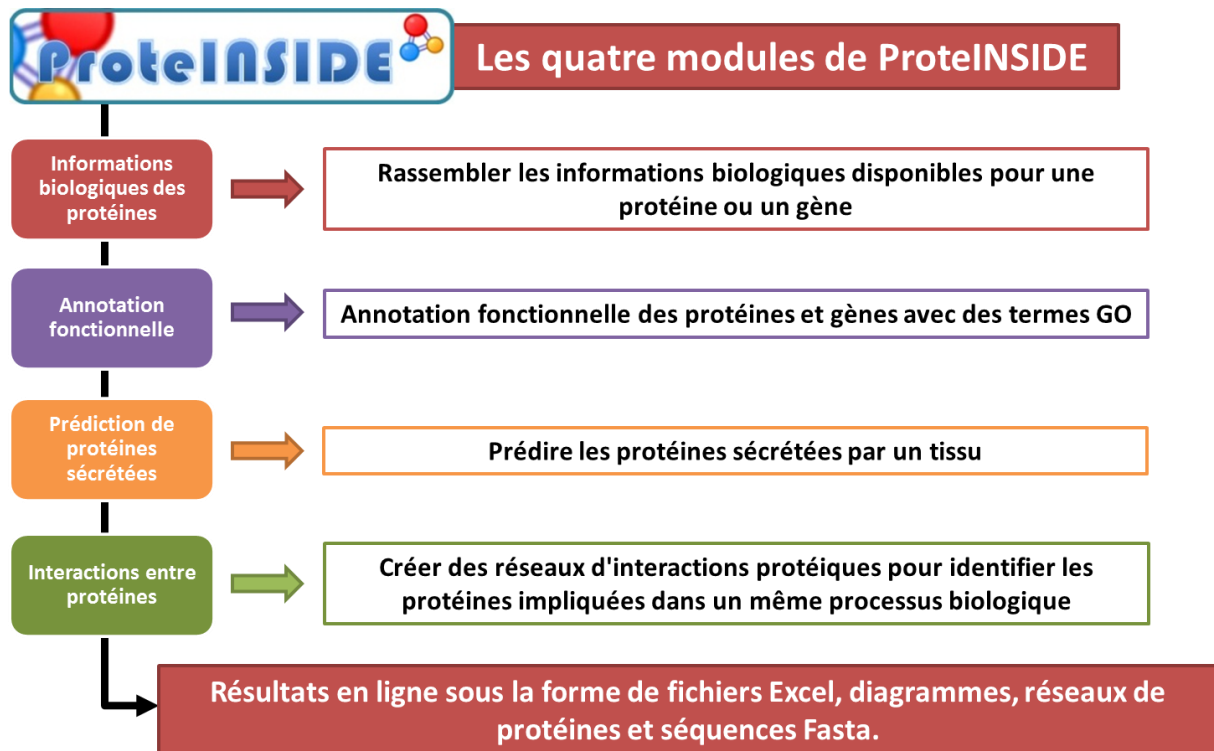


Figure 29 : Les quatre modules initiaux de ProteINSIDE.

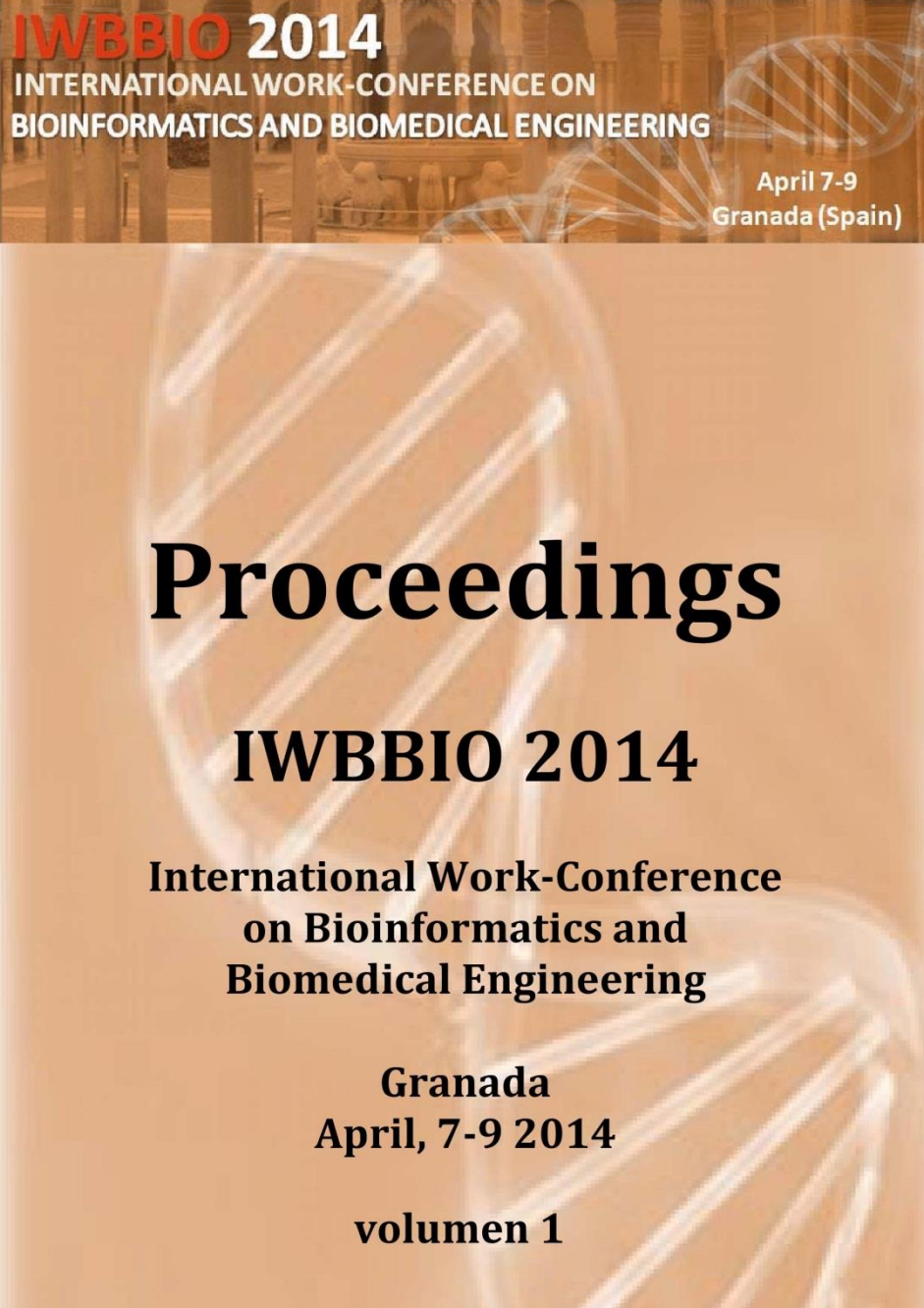
ARTICLE 1**Protein function easily investigated by genomics data mining using
the ProteINSIDE web service**

Nicolas KASPRIC^{1, 2}, Brigitte PICARD^{1, 2}, Matthieu REICHSTADT^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213, Lempdes, France

Proceedings IWBBIO 2014 (2th International Work-Conference on Bioinformatics and Biomedical Engineering) ; Grenade (Espagne) – Publié le 07 avril 2014 (12 pages).



Protein function easily investigated by genomics data mining using the ProteINSIDE web service

Nicolas KASPRIC^{1,2}, Brigitte PICARD^{1,2}, Matthieu REICHSTADT^{1,2}, Jérémy TOURNAYRE^{1,2} and Muriel BONNET^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Corresponding authors: nicolas.kaspric@clermont.inra.fr & muriel.bonnet@clermont.inra.fr

Abstract. ProteINSIDE is a new workflow to analyse lists of protein or gene identifiers from ruminant species and gather biological information provided by functional annotations, putative secretion of proteins and proteins interactions networks. ProteINSIDE gets results from several software and databases with a single query. From a unique list, ProteINSIDE uses orthologs identifiers within well studied species (Human, Rat or Mouse) to extend analyses and biological information retrieval. ProteINSIDE is freely available at: <http://www.proteinside.org>.

Keywords: web service, workflow, protein-protein interaction, protein secretion, gene ontology, networks.

1 Introduction

The current challenge for scientists working on the efficiency of ruminant (cattle, sheep or goat) and the quality of their products (meat, milk...) is to understand which genes and proteins control nutrient metabolism and partitioning between tissues or which genes and proteins control tissues growth and physiology [1]. Such questioning leads to the genome annotation, the sequencing and the quantification of gene expression or protein abundance. The quantity of data produced by these genomic and proteomic studies increases continuously [2-4]. There is a necessity to analyse, understand and generate biological information and knowledge from these data [5]. This is possible by using a panel of tools requiring different identifiers (IDs) per protein or gene and time to read and analyse the results. Moreover, most databases (DB) like UniProtKB [6] or NCBI [7] possess a large quantity of information and most of existing bioinformatic tools implemented as web services are specific to one analysis: as the annotation according to the Gene Ontology (GO) [8] or the prediction of signal peptide [9] or the molecular interactions identification [10] and visualization as networks [11, 12]. Many workflows that integrate several analyses are available [13-16]

and are specific to a species (*Drosophila*, *Arabidopsis thaliana*, *Escherichia coli*...), and thus are not suitable for the analysis of genomic data from ruminant species. The few workflows working with ruminant data are multispecies, the results are not species-specific and the data source is not available because of the privacy of databases (as the license software Pathway Studio [17]). Other workflows are specialized on the identification of candidate genes related to diseases as ToppGene [18]. Thus, to date there is no workflow dedicated to the integrative analysis of genomic data from ruminant species.

Unlike Human or model species like mouse or rat, ruminant species are less annotated and protein sequences are not always verified. Often, scientists use orthologs with the aim to increase the meaningful biological contexts for proteins. For this purpose biologists query for annotations according to Gene Ontology, the putative secretion of proteins, protein-protein interactions (PPI) and network analysis first in ruminant and then in Human or in rodents. The integration within a workflow of gateways between proteins / genes from ruminants and their orthologs from Human and models species has never been done.

Here we propose ProteINSIDE, a web service dedicated to a systematic and integrative analysis of protein's biological information. ProteINSIDE works using lists of proteins or genes IDs from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine) to annotate functions and cellular location, predict secreted proteins, search for interactions between proteins within and/or outside a dataset and allowing cross-species analysis using orthologs.

2 Materials and methods

This section lists necessary equipment, ProteINSIDE resources and describes the dataset used to assess the functionalities of our tool.

2.1 Equipment

ProteINSIDE doesn't require an installation on a computer and the web service is available online at www.proteinside.org by using an internet browser. ProteINSIDE is completely adapted for any internet browser, but for better performances we recommend to use Firefox, Chrome, or Safari.

2.2 Implementation

ProteINSIDE is divided into three parts: the workflow, the database and the web interface. The workflow is a combination of Perl (version 5.10.1; CPAN modules (Comprehensive Perl Archive Network) used and BioPerl [19]) and R (version 3.0.1; with "tnet" package [20]) scripts to query databases, recover protein data, perform calculations and run algorithms for signal peptide predictions and network visualisation. The MySQL database aims to reduce server load and thus stores both available knowledge from major public databases and results (and settings) from queries. The

web interface is programmed in PHP, HTML, and JavaScript. It is devoted to the creation of a new analysis, the view of results and users information with updates.

2.3 ProteINSIDE structure and interface

A flow chart (Fig 1) details the type of analysis (basic or customizable) and the four main queries proposed to the user. Whatever the type of analysis, the workflow uses data from the input file and runs default scripts (basic analysis) or scripts and options selected through the settings (customs analysis). At analysis completion, results are created and uploaded on ProteINSIDE database to decrease web interface treatment duration (results have to be deleted by the user; visitors results are automatically deleted monthly).

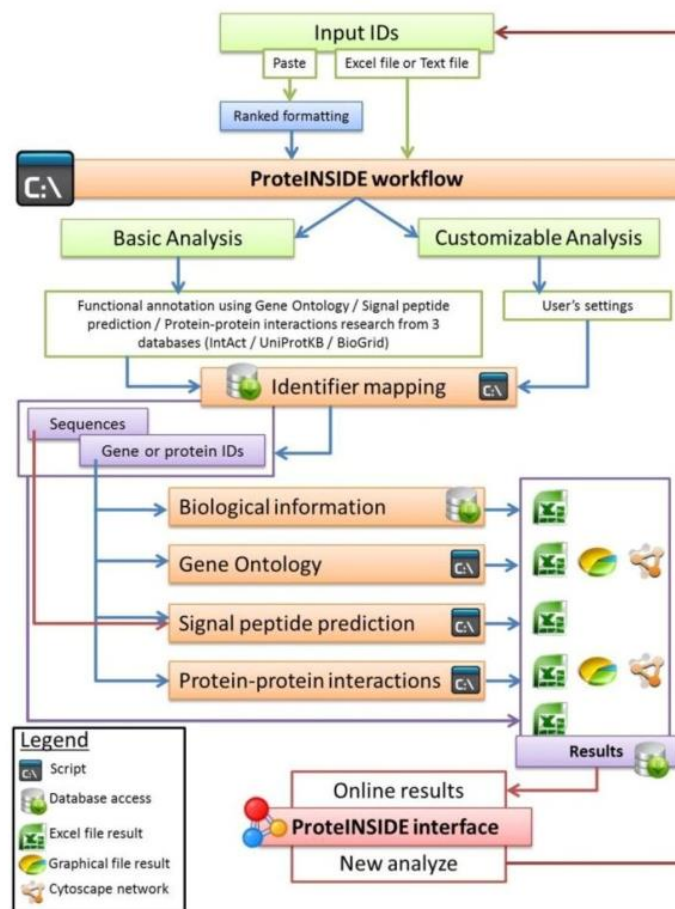


Fig. 1. Flow chart of ProteINSIDE structure. The four modules for querying the available biological information, annotations according to the gene ontology, signal peptide predictions and protein-protein interactions are either all present in the basic analysis or individually selected in the custom analysis.

ProteINSIDE is easily run by biologist through the interface. Registered user or visitor run a new analysis by using the web interface menus “Basic Analysis” (automatic settings) or “Custom Analysis” (user selects the settings):

1. Click on “Basic Analysis” menu on the homepage of ProteINSIDE
2. Fill in “the job name” box
3. Select the species for the analysis (related to the IDs that will be used on this query)
4. Upload your input file or directly paste your IDs
5. Click on the “Run the job” button to submit a new analysis

ProteINSIDE gives a link and an access code to view analysis status and get the results. The analysis status is indicated by the colour of a button: red for “analysis on the waiting list”, yellow for “the analysis is running” and green “analysis completion”. The blue globe is the link to access to the online results views:

1. Click on the blue globe button to view the results (use the trash to delete them)
2. Visualise the results summary produced by selected modules on the first default page
3. Navigate to module’s results pages by clicking on the module’s name on the toolbar menu.

2.4 The input and the output of ProteINSIDE

ProteINSIDE inputs are genes or proteins IDs (e.g. ADIPO or ADIPO_HUMAN) or UniProtKB protein accession numbers (e.g. Q15848). These IDs are uploaded as text tabulated files (extension .tab or .txt) or as Excel files (.xls or .xlsx). The input files have to be ranked as three columns (Fig. 2) because of the database format. Alternatively, the IDs are directly pasted.

	A	B	C
1	1	1	A4D1N9
2	2	2	ADIPO_HUMAN
3	3	3	C1T9A_HUMAN
4	4	4	F261_HUMAN
5	5	5	F262_HUMAN
6	6	6	PFKFB3
7	7	7	PFKFB4

Fig. 2. Example of an input files made using Excel 2010 and formatted for an upload.

The output files are Excel file (.xls), Cytoscape file (.cys or .xgmml), text or FASTA file (.txt or .fa) and pictures (.jpg or .png or .pdf). They are downloadable from the page results of each module of analysis.

2.5 The sample dataset

We created a dataset to assess ProteINSIDE performances. This dataset is composed of the UniProtKB accession numbers of 133 proteins (Table 1): 34 proteins related to the glycolysis cycle, 11 proteins from the respiratory chain, 5 proteins from the tricar-

boxylic acid cycle, 79 hormones or secreted proteins and proteins with very specific functions unrelated to the others. We also included a duplicated protein among proteins of the glycolysis to verify its recognition by ProteINSIDE.

We created this dataset on bovine species, but the numbers of annotations and PPI weren't sufficient for a clear representation of the functionalities of ProteINSIDE. Then, we used the same proteins in Human to test ProteINSIDE with the "Basic" and the "Custom Analysis" (Table 1).

Table 1. Results summary of ProteINSIDE analysis performances. The numbers are the proteins that belong to main pathways in the sample dataset, that are properly annotated by GO terms relevant to glycolysis and tricarboxylic acid (TCA) functions, and that have been predicted as secreted by SignalP for hormones.

Analyses and data	Glycolysis	Hormones	TCA	Analysis time (min)
Dataset	33+1 (duplicate)	79	5	-
Basic Analysis	29	78	3	2
Custom Analysis	33	78	5	10

3 Results and discussion

Here we present the results produced by a "Basic Analysis" and a "Custom Analysis" from our sample dataset, and we discuss the relevance of biological information extracted by ProteINSIDE. All of the 133 proteins were recognized by ProteINSIDE, the protein in duplicate was identified and excluded from the analysis (Fig. 3). Thus, 132 proteins were submitted to the analyses. The numbers of proteins / genes submitted to the analyses, numbers of annotations, PPI and predicted secreted proteins are recorded on the default page following the access to the results (Fig. 3).

General information	Analyze general results
<p>Name: ProtHumainglycolyseChaineResTCA Type: Basic Species: HUMAN Date: 2014-01-23</p> <p>Analysis parameters: ID Mapping / Gene Ontology / SignalP / Interaction research (IntAct / UniProtKB / BioGrid)</p>	<ul style="list-style-type: none"> ▣ Query find in ID Mapping DB: 133 ▣ Blastp alignment executed: 0 ▣ All GO detected: 568 for 123 annotated protein(s) ▣ Signal peptides detected: 85 ▣ Interactions detected on the dataset: 36
<p>Incomplet informations</p> <ul style="list-style-type: none"> ▣ Incomplet query: 12 + show/hide + ▣ Duplicate query: 1 + show/hide + 	

Fig.3. A table is provided by the default page after the access to the results both for a basic and a custom analysis. In addition to general information, the table provide counts of results retrieved by each module that has been run, incomplete query (IDs with missing biological information) and duplicate query.

3.1 Results of the Basic Analysis

The first module of analysis has extracted and summarized, as a downloadable table, other gene or protein IDs, gene or protein names, a summary for the protein function, the gene chromosomal location, information on tissue expression and cellular location, and the species in which orthologs have been identified. Thus on the “ID resume” page of the toolbar menu, a user has access at a glance to several information for a list of genes or their products, and also to direct links with the UniProtKB and the NCBI databases.

Multiple annotations GOs

No of annotated protein by GO	Function	Proteins	Gene Name	Number of items	Gene products detected on HUMAN			
7/0				10				
GO	Function	Proteins	Gene Name	Ontology group	Count	GO frequency (%)	Number of Gene products detected / expected	Percent of Gene products detected / expected
GO:0005096	glycolysis	Q01813 Q16877 P09104 P16119 P09338 P05558 P15259 P08237 P04496 P14618 P30613 P07295 P18969 P22739 P01174 P036811 Q05826 P04075 P19367 Q73366 Q16875 P09872 P06733 Q14566 P13929 P05062 P17858 P06744	PFKP PFKFB4 ENO2 PFKFB1 LDHA PGK1 PGAM2 PFKM GAPDH PKM PKLR PGK2 PGAM1 HK2 TR1 PGM1 PFKFB2 ALDOA HK1 ENTPD5 PFKFB3 ALDOC ENO1 GAPDH5 ENO3 ALDOB PFKL GPI	BP	28	22.764	28 /45	62.222
GO:0005184	neuropeptide hormone activity	P13689 P10092 P01282 P05850 P1309 P01185 P22466 P06307	UCH1 CALCB VIP CRH ADCYAP1 AVP GAL CCK	MF	8	6.504	8 /15	53.333
GO:0005179	hormone activity	P11148 P05850 P01298 P06307 P01277 P09529 P12272 P01242 P01282 P1278 Q9Y581 P04808 P01350 Q14405 Q95399 P01225 Q43555 P01189 P10882 P09681 Q15848 P01309 P09683 P01233 P52823 P08476 P01270 P05111 P16869 Q76961 P01216 P01222	GHRH1 CRH PPY CCK PRLH IHHB6 PTHLH GH2 VIP SST INSLS RLH1 GAST CSHL1 UTS2 FSHB GNRH2 POMC PPY GIP ADIPOQ INS SCT CGB8 STC1 IHHBA PTH IHHA NPPB STC2 CGA TSHB	MF	32	26.016	32 /62	51.613

Fig. 4. Results of the functional annotations according to the Gene Ontology are available as dynamic tables. Results can be sorted by: the GO's identifier, the function, protein ID or gene name, the ontology group, the number of annotated proteins or the number of expected gene products.

On the “GO” page of the toolbar menu, we checked the relevance of the annotations extracted by ProteINSIDE by looking for the over-representation of annotations relative to glycolysis and hormones. First, among the 132 proteins submitted, ProteINSIDE annotates 123 proteins with 568 unique GOs (Fig 3). We classed these GO according to the number of proteins annotated by GO and the percentage of gene products detected/expected to identify the most common pathways associated to our sample dataset (Fig. 4). By this way, we retrieved as the most common pathways: glycolysis and hormone activity (about 62% and 52% of expected annotated gene products with these GO in Human, respectively). We have to note a lack of annotations for 12 proteins of the sample dataset, and a lack of annotations relative to glycolysis for 4 proteins (28 of the 33 expected proteins related to the glycolysis were annotated; Table 1). This lack of annotations is related to our choice to use only GO terms that have been agreed by review curator in the “Basic Analysis”. This means that the “Basic Analysis” doesn't use GO annotations with IEA (Inferred by Electronic Annotation) evidence code, but the option to use IEA is provided in the custom analysis to extend the annotations.

On the “Secreted protein” page of the toolbar menu, the proteins potentially secreted are listed in a dynamic table (Fig. 3 and 5). From our sample dataset, 85 proteins were predicted as secreted by SignalP [9], among them 78 of the 79 proteins that were expected (Table 1). This lack of perfect prediction of the protein can be explained by the false positive and false negative prediction rates of SignalP, as already evaluated [21]. The prediction of secretion is then confirmed by a search for GOs related to the “secretion” function. Over the 85 predicted secreted proteins, 63 were annotated by GOs related to the “secretion” function. The double query of protein secretion both by the peptide signal prediction from protein sequence and the GO annotation is unique to ProteINSIDE.

Proteins	GO related to secretion	Gene Name	Number of rows	Signal peptides detected	Download table
			10	85 (on 133 proteins imported)	

Proteins	Protein ID	Gene Name	Peptide	GO related to secretion	Number of GO
Q9UBU3	GHRL_HUMAN	GHRL	noTM	GO:0051464 GO:0060124 GO:0005576 GO:0034774 GO:0005615 GO:0030252 GO:0051461 GO:0032024 GO:0043400	9
P01308	INS_HUMAN	INS	noTM	GO:0050796 GO:0005576 GO:0050715 GO:0034774 GO:0005615 GO:0090277 GO:0050708	7
Q15848	ADIPO_HUMAN	ADIPOQ	noTM	GO:0005576 GO:0005615 GO:0045715 GO:0034383	4
P01189	COLL_HUMAN	POMC	noTM	GO:0005576 GO:0034774 GO:0005615 GO:0030141	4
P08476	INHBA_HUMAN	INHBA	noTM	GO:0046881 GO:0005576 GO:0042701 GO:0046880	4
P16860	ANFB_HUMAN	NPPB	noTM	GO:0005576 GO:0005615 GO:0007589	3
P06850	CRF_HUMAN	CRH	noTM	GO:0051464 GO:0005576 GO:0005615	3
P09681	GIP_HUMAN	GIP	noTM	GO:0050796 GO:0005576 GO:0034774	3
P01275	GLUC_HUMAN	GCG	noTM	GO:0050796 GO:0005576 GO:0034774	3
P35318	ADML_HUMAN	ADM	noTM	GO:0005576 GO:0005615	2

Fig. 5. Results as a dynamic table, of the potentially secreted proteins predict by SignalP.

On the “Protein interactions” page of the toolbar menu, proteins within the dataset are linked by the “Interaction detection methods” or reviewed by a curator (clicking on node gives information about the protein and a link to UniProtKB database). We selected to query BioGrid [22], UniProtKB [6] and IntAct [23] because these PPI databases are reviewed by curators and the query of PPI in 2 or 3 PPI database delivered best results [24]. PPI are listed by a dynamic table or viewed as a network (Fig. 6). The interactions research between proteins of our sample dataset has identified 36 PPI that involved 23 different proteins. As expected, PPI within the sample dataset linked proteins known to contribute to the pyruvate dehydrogenase complex (Fig 6A), the complexes IV (Fig 6B) and I (Fig 6C) of the respiratory chain, and also some proteins linked to the glycolysis and the carbohydrate oxidation (Fig 6D and 6E).

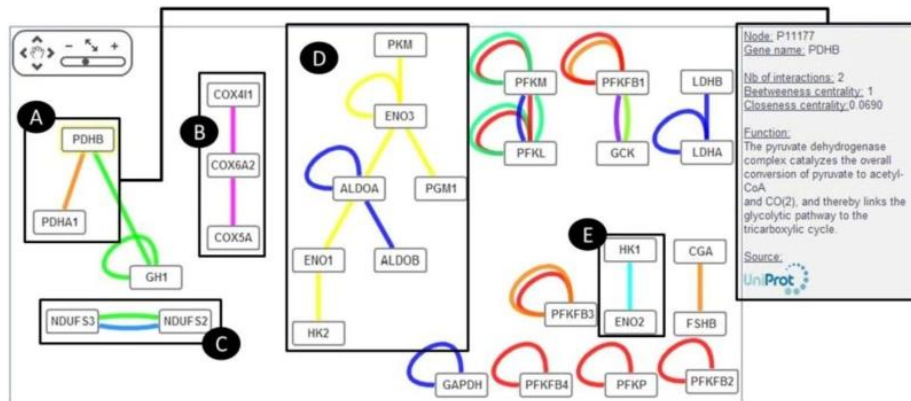


Fig. 6. Network of PPI retrieved by querying BioGrid, UniProtKB and IntAct databases. PPI are between proteins within the sample dataset. Information about a protein is obtained by clicking on a protein/gene or node. Edge colour depends on the detection method of the PPI.

3.2 The Custom Analysis: an added-value provided by the extension of the analysis

We made a “Custom Analysis” using the same major settings as the “Basic Analysis” but we used the proposed additional options:

1. The GO electronic annotation (IEA) evidence codes proposed to extend the annotation
2. GOTree network to view linked GOs
3. The search and the view of the PPI between proteins from our sample dataset and proteins outside the dataset (but still in the same species, here in Human) to extend the network and the biological information.

The use of electronic annotation has increased both the number of annotated proteins (132 rather than 123 without IEA in the basic analysis) and of annotations by around 50% since 1031 unique GOs were retrieved by ProteINSIDE. Over the 33 expected proteins related to the glycolysis, the Custom analysis of ProteINSIDE has annotated 32 proteins with the GO 0006096, glycolysis (Table 1). The GOTree network linked 236 GOs. We have chosen to visualize the GOs of the “Molecular Function” group (Fig. 7). In this visualisation, the dark red colour represents the most common GO associated to our sample dataset. As expected the GO:0005179, nominates “Hormone activity”, which is consistent with the over-representation of hormones in our sample dataset. This network has also linked more specific GOs or child terms [25] of the “Hormone activity” GO, as for example “Neuropeptide hormone activity” (GO:0005184).

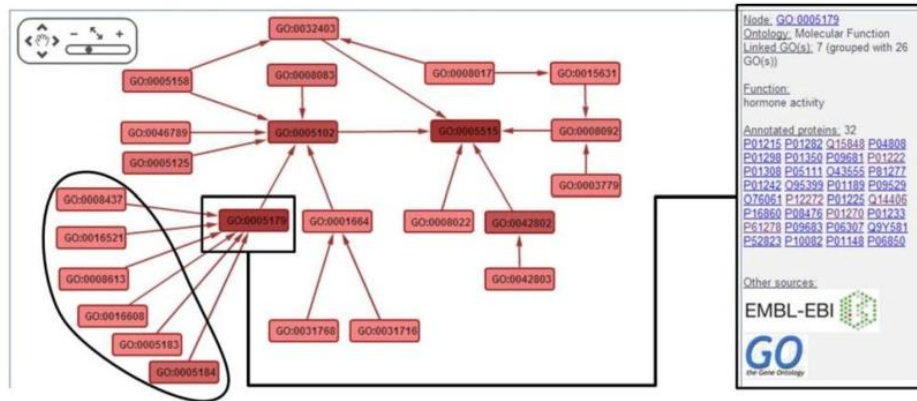


Fig. 7. A Network that links GOs used to annotate proteins of the sample dataset. Red colour is only for the GO terms relative to the Molecular Function. The degree of colour saturation represents the quantity of proteins annotated by a GO. Each edge means that a term A is a sub-type of a term B (is_a). Information about a GO is obtained by clicking on the GO or the node.

Lastly, the same proteins as the “Basic Analysis” were predicted as secreted (Table 1). Thanks to the IEA electronic annotation, 82 proteins over the 85 proteins predicted to be secreted by SignalP, were also annotated by GOs related to the “secretion” function. It’s 19 more than the 63 of the “Basic Analysis” because of the use of IEA evidence code for GO annotation.

By comparison with “Basic Analysis”, “Custom Analysis” searches for PPI between the proteins within and outside the sample dataset by querying up to 28 DB. We chose to query the same 3 DB (BioGrid, UniProtKB and IntAct) to compare with the “Basic Analysis”. ProteINSIDE retrieved 616 PPI made by 221 proteins among them 61 from the dataset. We visualized the network of the PPI (Fig. 8) and we retrieved some sub-networks relative to the respiratory chain (Fig. 8A), hormone activity such as signalization by adipokines (Fig. 8B), the growth hormone (Fig. 8C) and thyroid hormones (Fig. 8D), as well as sub-networks relative to glycolysis and carbohydrate metabolism (not highlighted).

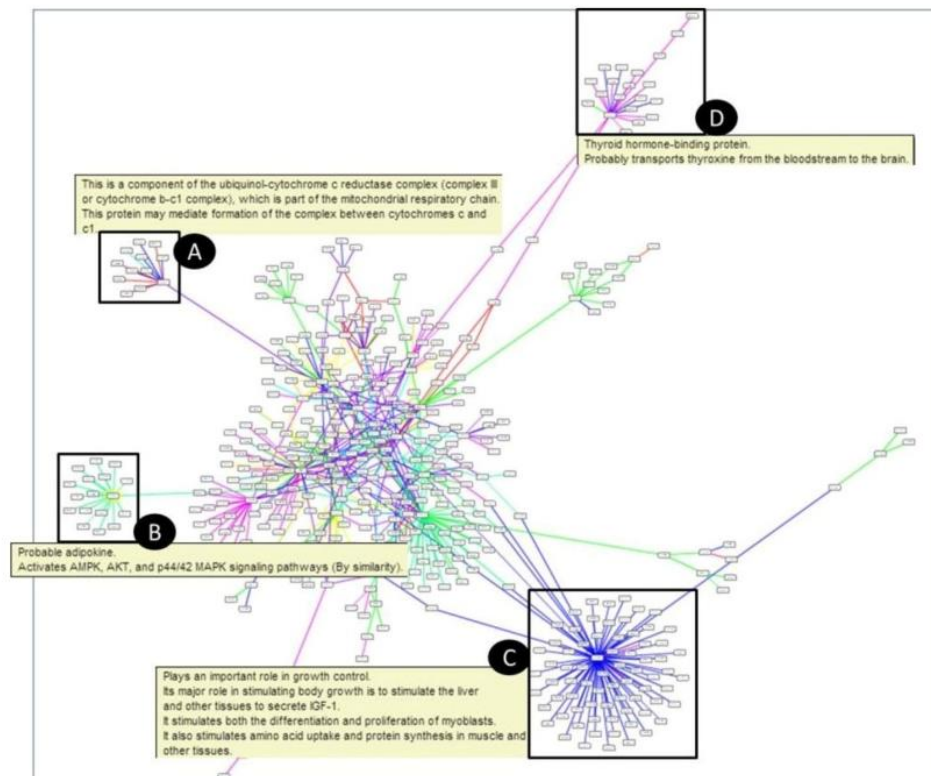


Fig. 8. Network of PPI retrieved by querying the BioGrid, UniProtKB and IntAct DB and using PPI with human proteins outside of the dataset.

4 Conclusion

In this work we present the performances of ProteINSIDE, a new powerful workflow which gathers tools and public databases to retrieve biological information of genes or proteins lists from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). The presented web service has correctly identified a dataset of 133 proteins, has excluded a duplicate query and has retrieved biological information for each protein. According to our dataset, ProteINSIDE properly annotates the proteins related to the glycolysis, the proteins affiliated as hormones, and the putatively secreted proteins. ProteINSIDE has revealed the most common pathways related to our dataset by creating networks from PPI interactions within and outside the dataset and from links between GOs. Each result is easily accessible and downloadable.

ProteINSIDE offers a great support to analyse a large quantity of data from genomic and proteomic studies. ProteINSIDE is also the unique web service that makes all of these analyse using ruminant IDs.

Acknowledgment

This work was supported by the Region Auvergne (FRANCE) and Apis-Gene (FRANCE).

References

1. Bonnet M, Cassar-Malek I, Chilliard Y, Picard B. "Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species". *Animal : an international journal of animal bioscience*, 4(7), pp. 1093-109. PubMed PMID: 22444612. (2010)
2. Picard B, Cassar-Malek I, Guillemain N, Bonnet M. Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: Moo-Young M, editor. *Comprehensive Biotechnology (Second Edition)*. Burlington: Academic Press; 2011. p. 395-405.
3. Chaze T, Meunier B, Chambon C, Jurie C, Picard B. "Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle". *Animal : an international journal of animal bioscience*, 3(7), pp. 980-1000. PubMed PMID: 22444818. (2009)
4. Taga H, Chilliard Y, Meunier B, Chambon C, Picard B, Zingaretti MC, et al. "Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown?". *Journal of cellular physiology*, 227(4), pp. 1688-700. PubMed PMID: 21678425. (2012)
5. Woelders H, Te Pas MF, Bannink A, Veerkamp RF, Smits MA. "Systems biology in animal sciences". *Animal : an international journal of animal bioscience*, 5(7), pp. 1036-47. PubMed PMID: 22440099. (2011)
6. Magrane M, Consortium U. "UniProt Knowledgebase: a hub of integrated protein data". *Database : the journal of biological databases and curation*, 2011, pp. bar009. PubMed PMID: 21447597. Pubmed Central PMCID: 3070428. (2011)
7. Coordinators NR. "Database resources of the National Center for Biotechnology Information". *Nucleic Acids Res*, 41(Database issue), pp. D8-D20. PubMed PMID: 23193264. Pubmed Central PMCID: 3531099. (2013)
8. Binns D, Dummer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. "QuickGO: a web-based tool for Gene Ontology searching". *Bioinformatics*, 25(22), pp. 3045-6. PubMed PMID: 19744993. Pubmed Central PMCID: 2773257. (2009)
9. Petersen TN, Brunak S, von Heijne G, Nielsen H. "SignalP 4.0: discriminating signal peptides from transmembrane regions". *Nature methods*, 8(10), pp. 785-6. PubMed PMID: 21959131. (2011)
10. Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, et al. "PSICQUIC and PSISCORE: accessing and scoring molecular interactions". *Nature methods*, 8(7), pp. 528-9. PubMed PMID: 21716279. Pubmed Central PMCID: 3246345. (2011)

11. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. "Cytoscape 2.8: new features for data integration and network visualization". *Bioinformatics*, 27(3), pp. 431-2. PubMed PMID: 21149340. Pubmed Central PMCID: 3031041. (2011)
12. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. "Cytoscape Web: an interactive web-based network browser". *Bioinformatics*, 26(18), pp. 2347-8. PubMed PMID: 20656902. Pubmed Central PMCID: 2935447. (2010)
13. Wong AK, Park CY, Greene CS, Bongo LA, Guan Y, Troyanskaya OG. "IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks". *Nucleic Acids Res*, 40(Web Server issue), pp. W484-90. PubMed PMID: 22684505. Pubmed Central PMCID: 3394282. Epub 2012/06/12. eng. (2012)
14. Pache RA, Ceol A, Aloy P. "NetAligner--a network alignment server to compare complexes, pathways and whole interactomes". *Nucleic Acids Res*, 40(Web Server issue), pp. W157-61. PubMed PMID: 22618871. Pubmed Central PMCID: 3394252. Epub 2012/05/24. eng. (2012)
15. Renaud Y, Baillif A, Perez JB, Agier M, Mephu Nguifo E, Mirouse V. "DroPNet: a web portal for integrated analysis of Drosophila protein-protein interaction networks". *Nucleic Acids Res*, 40(Web Server issue), pp. W134-9. PubMed PMID: 22641854. Pubmed Central PMCID: 3394298. Epub 2012/05/30. eng. (2012)
16. Tuncbag N, McCallum S, Huang SS, Fraenkel E. "SteinerNet: a web server for integrating 'omic' data to discover hidden components of response pathways". *Nucleic Acids Res*, 40(Web Server issue), pp. W505-9. PubMed PMID: 22638579. Pubmed Central PMCID: 3394335. Epub 2012/05/29. eng. (2012)
17. Nikitin A, Egorov S, Daraselina N, Mazo I. "Pathway studio--the analysis and navigation of molecular networks". *Bioinformatics*, 19(16), pp. 2155-7. PubMed PMID: 14594725. (2003)
18. Chen J, Bardes EE, Aronow BJ, Jegga AG. "ToppGene Suite for gene list enrichment analysis and candidate gene prioritization". *Nucleic Acids Res*, 37(Web Server issue), pp. W305-11. PubMed PMID: 19465376. Pubmed Central PMCID: 2703978. (2009)
19. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. "The Bioperl toolkit: Perl modules for the life sciences". *Genome research*, 12(10), pp. 1611-8. PubMed PMID: 12368254. Pubmed Central PMCID: 187536. (2002)
20. T. O. "Structure and Evolution of Weighted Networks". University of London (Queen Mary College), London, UK, pp. 104-22. (2009)
21. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. "Locating proteins in the cell using TargetP, SignalP and related tools". *Nat Protoc*, 2(4), pp. 953-71. PubMed PMID: 17446895. (2007)
22. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. "The BioGRID interaction database: 2013 update". *Nucleic Acids Res*, 41(Database issue), pp. D816-23. PubMed PMID: 23203989. Pubmed Central PMCID: 3531226. (2013)
23. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. "The IntAct molecular interaction database in 2012". *Nucleic Acids Res*, 40(Database

- issue), pp. D841-6. PubMed PMID: 22121220. Pubmed Central PMCID: 3245075. (2012)
24. Martha VS, Liu Z, Guo L, Su Z, Ye Y, Fang H, et al. "Constructing a robust protein-protein interaction network by integrating multiple public databases". *BMC Bioinformatics*, 12 Suppl 10, pp. S7. PubMed PMID: 22165958. Pubmed Central PMCID: 3236850. (2011)
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium". *Nature genetics*, 25(1), pp. 25-9. PubMed PMID: 10802651. Pubmed Central PMCID: 3037419. (2000)

Conclusion de l'étude

La présentation de ProteINSIDE à un public de bioinformaticiens, d'informaticiens et de biologistes, notamment à la conférence IWBBIO, a révélé différents problèmes de conception, de performance et de validation des résultats. La première suggestion pour améliorer ProteINSIDE était d'intégrer un calcul d'enrichissement des termes GO pour représenter les annotations significativement surreprésentées dans un jeu de données. Une autre suggestion était de valider les prédictions de peptides signaux effectuées par SignalP, grâce à d'autres outils. Enfin, est apparue la nécessité d'ajouter une fonction pour trier les réseaux de PPI afin d'identifier les protéines centrales. L'objectif était à la fois de mettre en évidence les protéines qui interagissent pour une fonction biologique (réseaux entre protéines d'un jeu de données), mais aussi de révéler les protéines centrales dans un grand réseau (l'absence de cette protéine modifie fortement le réseau).

1.3 – ProteINSIDE : mises à jour pour renforcer la validité des résultats

Les retours d'expérience, des utilisateurs et des programmeurs, nous ont conduits à mettre à jour ProteINSIDE afin d'améliorer les modules d'analyses, l'affichage en ligne des résultats, et de corriger quelques failles.

Pour ce qui concerne l'affichage du service web et/ou des résultats, l'interface web de ProteINSIDE a reçu une homogénéisation des couleurs et des tableaux dynamiques ont été appliqués à tout le site (en utilisant la technologie Google Charts ; <https://developers.google.com/chart/>). L'objectif était de permettre à l'utilisateur de trier les tableaux de résultats en fonction de plusieurs critères. Par exemple, pour les résultats de GO, l'utilisateur peut désormais rechercher un terme GO particulier ou une protéine ou un gène, il peut sélectionner les termes GO en fonction de leur groupe d'ontologie. Enfin, il peut taper le début d'une fonction biologique pour exclure les termes GO n'ayant pas ce motif dans leur fonction. La colorisation de l'interface permet de garder un thème d'affichage identique pour toutes les pages de ProteINSIDE (interface utilisateur, tutoriaux, résultats...). Les formulaires de création des analyses PPI ont reçu des indicateurs de disponibilité du service Psiquic. En effet, ce service est parfois indisponible et les résultats des analyses sont faussés : manque de résultats ou aucun résultat. ProteINSIDE prend en compte cette indisponibilité et met en garde l'utilisateur quant au manque de résultats générés. Si une base de données indisponible est malgré tout sélectionnée, ProteINSIDE ne l'interrogera pas. La page de création des analyses à paramétrer a été refaite avec des boutons qui génèrent automatiquement les analyses les plus courantes (analyse basique avec interactions enrichies, Gene Ontology complète avec annotations électroniques, ou encore recherche de PPI entre des protéines du jeu de données et des protéines extérieures au jeu parmi les 32 bases de Psiquic). Une description complète de toutes les bases d'interactions proposées par Psiquic est regroupée dans un tableau afin que l'utilisateur puisse les sélectionner en fonction : des types d'interactions moléculaires (protéines – protéines, protéines – petites molécules et protéine – acide nucléique) et des données (prédites, revue par des experts ou appartenant aux consortiums IMEx et/ou MIMix). ProteINSIDE a aussi migré vers un serveur plus puissant (64 cœurs et 128 Go de mémoire vive). Malgré ce nouvel apport de ressources, le chargement des pages pour les analyses avec de nombreux ID était toujours trop long (plusieurs minutes pour 4000 ID chargés). Aussi, un pré-calcul (résultats auto-générés sans utilisation de la base de données de ProteINSIDE) des pages de résultats a été mis en place et a été réalisé par Antoine

Perot (stage IUT bioinformatique ; co-encadrement entre Matthieu Reichstadt, Jérémy Tournayre et moi-même).

Les modifications du workflow ont concerné les modules d'analyse « Gene Ontology », « Secreted protein » et « Protein Interaction ». Pour le module « Gene Ontology » le calcul de l'enrichissement des termes GO relatifs à une liste de gènes ou de protéines a été ajouté. Précédemment, le module de GO effectuait deux calculs : la fréquence d'apparition du terme GO dans le jeu de données (recouvrement du terme GO) et le pourcentage de protéines du jeu de données annotées par rapport à toutes les protéines annotées par ce terme GO dans l'espèce d'analyse. L'implémentation d'un test exact de Fisher, initialement publié dans l'outil FatiGO+ ([Al-Shahrour et al., 2007](#)), a été intégré à ProteINSIDE et met en évidence les termes GO (fonctions biologiques) surabondants. Le test calcule, à l'aide d'un tableau de contingence deux à deux, si l'annotation du terme GO est significativement surreprésentée dans le jeu de données par rapport à l'annotation de ce terme dans l'espèce d'étude. Ce test est réalisé grâce à un algorithme sous R et il est corrigé par la méthode de Benjamini & Hochberg ([Benjamini et Hochberg, 1995](#)). Ainsi, dans la version actuelle de ProteINSIDE, les termes GO obtenus lors d'une analyse sont triés en fonction de leur p -value résultant du test exact de Fisher et de sa correction (un code couleur met en évidence les termes les plus enrichis). Les termes GO sont triés en deux tableaux : les termes annotant un seul ID et les termes annotant plusieurs ID. Cependant, les utilisateurs souhaitant rechercher des termes ou des ID, dans la totalité des résultats de GO, peuvent utiliser un tableau supplémentaire regroupant tous les résultats. Le classement des termes GO selon leur p -value (ainsi que le code couleur associé à la p -value) a aussi été ajouté à la page de visualisation des termes sous la forme de réseau d'arbres de phylogénie.

Pour le module « Secreted proteins », deux nouveaux résultats appuient les prédictions de protéines potentiellement sécrétées par un peptide signal : la localisation cellulaire des protéines renseignées sur UniProt et une prédiction de cette localisation cellulaire par le logiciel TargetP ([Emanuelsson et al., 2000](#)). Des scores, calculés par les deux outils SignalP et TargetP, ont aussi été ajoutés pour informer les utilisateurs sur la qualité de la prédiction. En outre, pour identifier les protéines sécrétées par des voies de sécrétion qui n'impliquent pas de peptide signal (([Nickel, 2003](#)) cf. Étude bibliographique, paragraphe 6.3), nous avons ajouté la prédiction de la localisation cellulaire avec TargetP, et rapporté les GO de sécrétion qui annotent cette protéine. Pour appuyer les prédictions de sécrétion, le dictionnaire de termes GO liés à une fonction de sécrétion a été revu. Il contient désormais environ 1000 termes GO associés à la sécrétion comparativement aux 2000 termes utilisés dans la version bêta. Certains termes GO de la version bêta étaient sélectionnés alors qu'ils n'appuyaient pas la sécrétion potentielle d'une protéine. Par exemple, les termes encadrés en vert dans la [Figure 30 \(A\)](#) ne sont pas reliés à la sécrétion d'une protéine, mais participent à la création d'une structure cellulaire. Ils étaient sélectionnés en fonction des mots composant leur description (soulignés en rouge dans la [Figure 30](#)). Désormais, le dictionnaire est créé en parcourant toute l'arborescence des termes GO et en sélectionnant ceux qui ont un lien avec des fonctions de sécrétion plus hautes dans l'arborescence ([Figure 30 - B](#)).

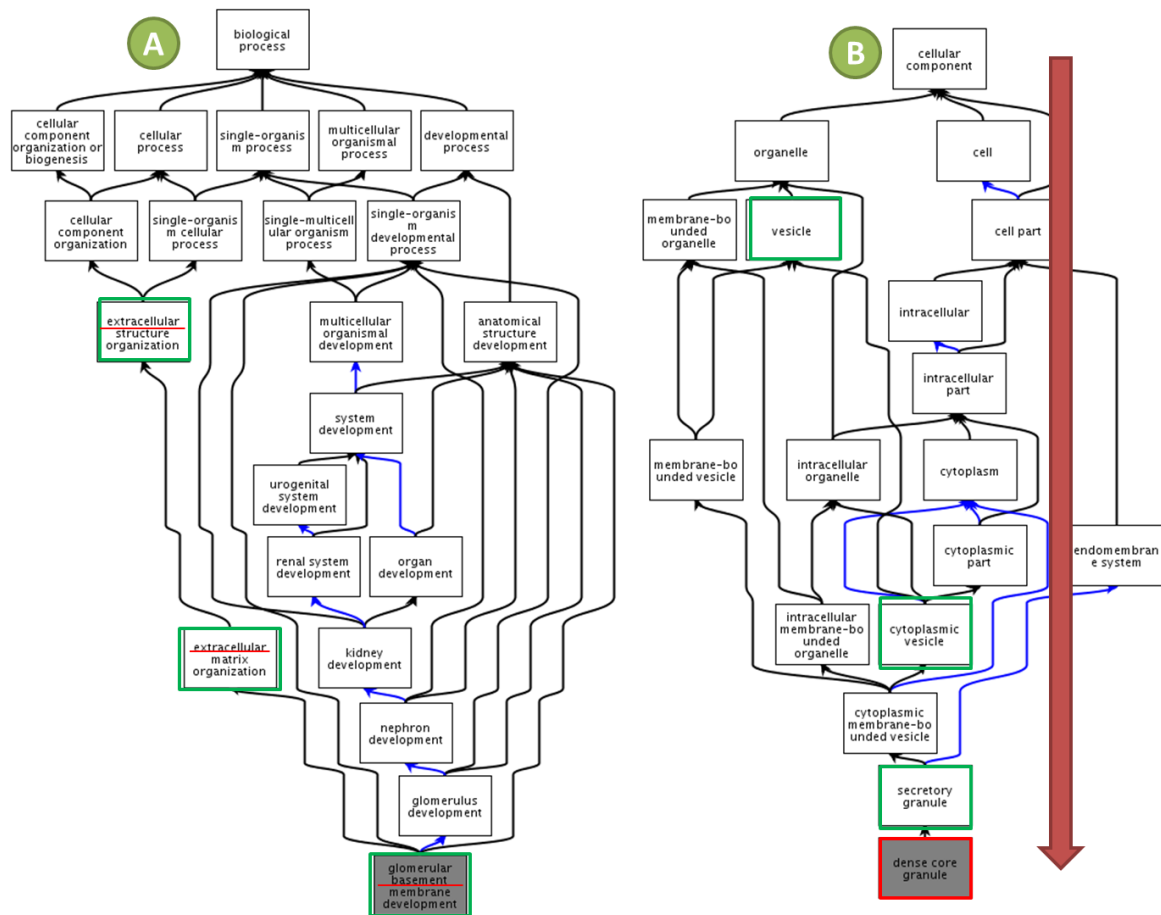


Figure 30 : Les méthodes de sélection des termes GO dans la version bêta et la version aboutie de ProteINSIDE sont comparées. (A) dans la version bêta, les termes GO reliés à l'état de sécrétion d'une protéine n'étaient pas forcément reliés à la sécrétion de protéines (termes encadrés en vert dans A). (B) les termes sont sélectionnés en partant du début de l'arborescence (encadrés en vert dans B), ainsi les termes qui n'étaient pas sélectionnés avant (terme en rouge dans B) sont pris en compte avec cette nouvelle méthode.

Le module PPI a été nettement amélioré. Un dictionnaire des méthodes de détection expérimentales a été ajouté à la base de données de ProteINSIDE afin de trier plus rapidement les résultats de Psicquic. Pour appuyer la robustesse d'une interaction entre deux protéines, le nombre de fois qu'une interaction est détectée par une méthode de détection est compté (Martha *et al.*, 2011). Plus une interaction a été détectée avec une ou plusieurs méthode(s), plus elle est fiable. Des améliorations visuelles au niveau des réseaux d'interactions ont été ajoutées. Les protéines prédites comme sécrétées par SignalP (si le module est activé) sont notées sur le réseau (Figure 31). Il est possible d'afficher uniquement un groupe de nœuds en sélectionnant les protéines souhaitées. Certaines interactions entre deux protéines sont prouvées par des dizaines de méthodes de détection, le réseau peut être surchargé, ainsi un bouton permet de n'afficher qu'un seul lien entre chaque protéine (une seule méthode de détection ; Figure 31). Enfin, il est possible de réaliser une nouvelle analyse sur un groupe de nœuds, notamment pour effectuer une annotation par des termes GO. Nous avons aussi ajouté les méthodes statistiques de « betweenness » (intermédiaire) et « closeness centrality » (proximité) pour trier les nœuds des réseaux (Hwang *et al.*, 2008). Ces deux

méthodes statistiques mettent en évidence des protéines clés dans un réseau en fonction des centralités :

- intermédiaires, qui quantifient la fréquence d'un nœud dans les plus courts chemins entre chaque paire de nœuds afin de détecter « les goulots d'étranglement » dans un réseau
- de proximités, qui quantifient les plus courts chemins à partir d'un nœud donné vers tous les autres nœuds, une grande proximité indique que le nœud est proche du centre topologique du réseau

Ces informations statistiques sont aussi disponibles sur les réseaux téléchargés à partir des fichiers de ProteINSIDE.

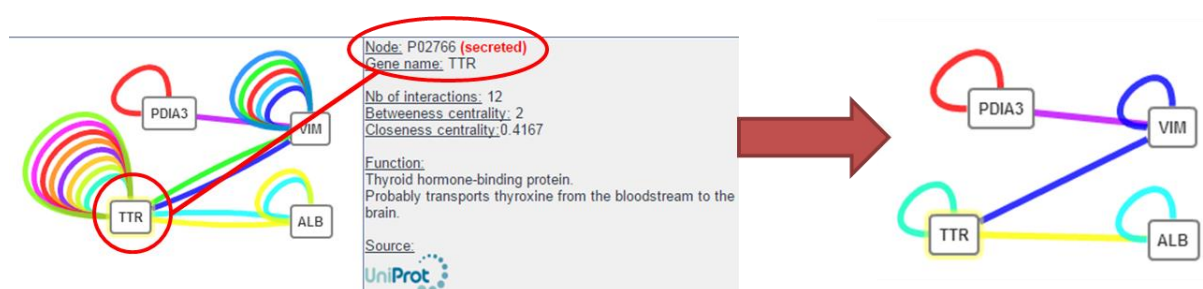


Figure 31 : Un clic sur une protéine donne des informations la concernant via la fiche UniProt et la prédiction de SignalP (si le module est utilisé). L'affichage d'un seul lien entre deux protéines améliore considérablement l'affichage.

À chaque analyse, un fichier qui rassemble les résultats de tous les modules utilisés est désormais créé automatiquement au format Excel. Il stocke tous les résultats sans avoir à parcourir chaque page du site pour les télécharger manuellement (les séquences au format Fasta ne sont pas intégrées au fichier). D'autres mises à jour invisibles aux utilisateurs ont été réalisées, notamment des améliorations pour la reconnaissance des ID des jeux de données et les homologies entre espèces. Les différentes bases de données biologiques (termes GO liés à la sécrétion, connaissances biologiques et annotations) sont désormais mises à jour de façon automatique chaque mois. La nouvelle version de ProteINSIDE et un tutoriel pour son utilisation et la visualisation des résultats ont été publiés dans Genomics and Computational Biology.

ARTICLE 2**Protein function easily investigated by genomics data mining using
the ProteINSIDE online tool**

Nicolas KASPRIC^{1, 2}, Matthieu REICHSTADT^{1, 2}, Brigitte PICARD^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213, Lempdes, France

Genomics and Computational Biology, doi: 10.18547/gcb.2015.vol1.iss1.e16. – Publié le 06 septembre 2015 (14 pages).

Resource

Protein Function Easily Investigated by Genomics Data Mining Using the ProteINSIDE Online Tool

Nicolas KASPRIC^{1,2,*}, Matthieu REICHSTADT^{1,2}, Brigitte PICARD^{1,2}, Jérémy TOURNAYRE^{1,2} and Muriel BONNET^{1,2,*}

¹INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

²Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

*To whom correspondence should be addressed: nicolas.kaspric@clermont.inra.fr; muriel.bonnet@clermont.inra.fr

Received 2015-02-13; Accepted 2015-06-09; Published 2015-09-18

ABSTRACT

Nowadays, genomic and proteomic studies produce vast amounts of data. To get the biological meaning of these data and to generate testable new hypothesis, scientists must use several tools often not designed for ruminant studies. Here we present ProteINSIDE: an online tool to analyse lists of protein or gene identifiers from well-annotated species (human, rat, and mouse) and ruminants (cow, sheep, and goat). The aims of ProteINSIDE modules are to gather biological information stores in well-updated public databases, to proceed to annotations according to the Gene Ontology consortium, to predict potentially secreted proteins, and to search for proteins interactions. ProteINSIDE provides results from several software and databases in a single query. From a list of identifiers, ProteINSIDE uses orthologs or homologs to extend analyses and biological information retrieval. As a tutorial, we presented how to launch, to recover, to view, and to interpret the results provided by the two types of analysis available with ProteINSIDE (basic and custom analyses). ProteINSIDE is freely available using an internet browser at www.proteinside.org. The results of this article are provided on the home page of ProteINSIDE website as the example of an analysis result.

AVAILABILITY AND REQUIREMENTS

- Project name: ProteINSIDE
- Project home page: <http://www.proteinside.org>
- Operating systems: Linux, MacOS, Windows
- License: No

KEYWORDS

Online tool, workflow, protein function, protein interaction, protein secretion, gene ontology, networks, ruminant, genomics.

INTRODUCTION

Given the increasing amount of genomic and proteomic data produced even in ruminants [1, 2, 3], there is a challenge for the bioinformatic data processing, which has not yet been completely solved

[4]. Such bioinformatic data processing has to proceed to data gathering and database searching in order to produce a functional interpretation of large datasets. For this purpose, workflows integrating several bioinformatics analyses are now available [5-8] and were developed to mine dataset from specific species (BioMyn [9] for human, DroPNet [7] for *Drosophila*, TAIR [10] for *Arabidopsis thaliana*, EcoCyc [11] for *Escherichia coli* ...) or to identify candidate genes related to diseases as ToppGene [12] or NetPath [13]. The few workflows currently used for the bioinformatics data processing of ruminant datasets are multispecies. Consequently, the data source of the results proposed is not available because of the privacy of databases (as the licensed software Pathway Studio [14] or Ingenuity Pathway Analysis (www.ingenuity.com; Redwood City, CA, USA). An alternative for scientists working with ruminant datasets is to use dedicated and complementary bioinformatics tools implemented as web services. These tools are dedicated to one type of analysis, as for example the annotation according to the Gene Ontology (GO) [15], the prediction of signal peptide to identify putative secreted proteins [16], or the molecular interactions identification [17] and visualization as networks [18, 19]. Whatever the analysis carried out, the first step is to connect a protein name to a unique identifier (ID). Conversely to gene names that have been standardized, protein names or IDs can differ between databases or tools, especially for ruminant data that remains to be largely curated in most of databases. Thus, the use of several bioinformatics tools to mine ruminant datasets leads to a substantial loss of information and time.

A strategy to perform a systematic and integrative analysis of biological protein information from ruminant datasets is to develop an online workflow that integrates several analysis steps in one package and from a unique ID. Thus, we propose ProteINSIDE [20], a web service dedicated to a systematic and integrative analysis of protein's biological information from ruminant datasets. Unlike human, mouse or rat, ruminant species are less annotated and protein sequences or information are not always curated. Often, scientists working with ruminant use orthologs or homologs with the aim to increase the meaningful

biological contexts for proteins thanks to knowledge available in well-annotated species. Thus, ProteINSIDE was designed to run using lists of protein or gene IDs from 6 species (cow, sheep, goat, human, rat, and mouse) to annotate biological and molecular functions and cellular location, predict secreted proteins, search for interactions between proteins within and/or outside a dataset. The objective of this article is to propose a tutorial to use ProteINSIDE and interpret generated results.

METHODS

This section lists necessary equipment, ProteINSIDE resources and describes the dataset used to assess the functionalities of the tool.

ProteINSIDE's features

ProteINSIDE is an online workflow with an interface devoted to accessible and fully customisable analyses from lists of protein or gene IDs. Registered users have access to an analyses manager to run and save analysis, and visualise the results. Unregistered users can use ProteINSIDE, but there is no analyses manager and analyses are deleted each month. ProteINSIDE is divided into three parts: the workflow, the database, and the web interface (Figure 1).

The web interface, designed to easily use ProteINSIDE, helps the user to create the analyses, to have access to the results thanks to a balance between technical functionalities and visual elements, and to inform about updates (Figure 1). ProteINSIDE proposed two types of analysis to be launched: the basic analysis (automatic settings) and the custom analysis (user's settings). There is also a pre-set analysis for registered users only who want to make a new analysis with settings of a previous analysis.

The basic analysis performs a:

- Functional annotation using GO terms by querying QuickGO database [21] without electronic annotation.
- Prediction of secreted proteins using SignalP [16] and TargetP [22] software. We improve the prediction by giving GO terms related to the

cellular location of the protein and the processes of secretion.

- Search of proteins interactions curated and listed in IntAct [23], UniProt [24], and BioGrid [25] databases.

The custom analysis performs programs and their options that have to be selected by the user in order to:

- Perform a functional annotation using GO terms from QuickGO, with the options to select also electronic annotations (predicted and scripted annotations), and to generate a GOTree view of linked GO terms (pathways of functional annotation).
- Predict secreted proteins with the option to increase the software's sensitivity of prediction and by this way to increase the number of predicted proteins, however with a higher number of false positive results.
- Search for protein interaction within (core network) and outside (extended network) the uploaded dataset. Options propose to select interactions stored within 1 to 31 databases gathered by the PSICQUIC website [17]. User can select the databases depending on the type of interactions (PPI, Nucleic acid-Protein interaction (NPI), and Smallmolecule-Protein interaction (SPI)) and the data (curated, predicted, curated according to the IMEx project [26] or the MIMix curation [27]). PSICQUIC service or some databases could be offline, that's why the status of each website is indicated in the table.

To submit an analysis, users either directly paste a list of IDs or upload a file of IDs. Inputs can be protein (e.g., ADIPO_HUMAN) or gene (e.g., ADIPO or gil62022275) ID, or protein accession numbers (e.g., Q15848) from six species: cow, human, rat, mouse, sheep, and goat (Figure 2). A new analysis is run directly or is placed on a waiting list if the workflow is overloaded. Uploaded data and results remain confidential.

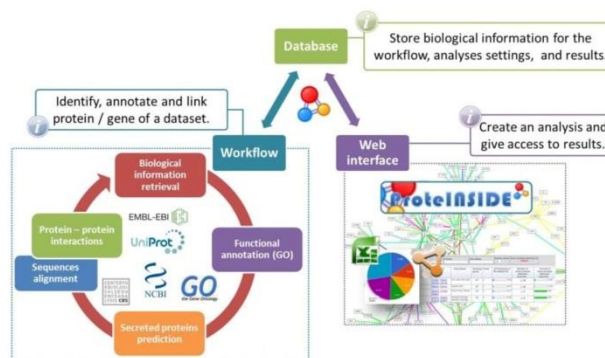


Figure 1. Flow chart of ProteINSIDE structure. The four workflow's modules are either all launched in the basic analysis or individually selected in the custom analysis. These modules aims to query the available biological information, annotate according to the gene ontology, predict signal peptide and visualized protein-protein interactions.

In addition to the web interface, ProteINSIDE is composed of a database and a workflow.

The database (invisible to any user) collects and stores the information required for the proper functioning of ProteINSIDE. It stores analysis settings and results to reduce server load (Figure 1). The database stores also a gathering of biological information from the NCBI [28] (Gene, Protein, and HomoloGene for known orthologous proteins between the 6 species) and UniProt [24] databases (for the ID Mapping module), and QuickGO [21] and AmiGO [29] (for the GO annotation module). A script updates automatically and monthly the database by extracting IDs, homologs, biological function, FASTA sequence, and other information from the latest releases of these databases.

The workflow uses uploaded data. It is a combination of Perl and R scripts to query databases, recover protein data, perform calculations and run algorithms for signal peptide predictions and network visualisation (Figure 1). The workflow is invisible to any user. The workflow is composed of 4 parts: the "ID Mapping", the search of annotations according to GO, the prediction of secreted proteins, and the search of protein-protein interactions (PPI). The workflow always starts by the ID Mapping program which searches the biological information available for each protein or gene of the input within the ProteINSIDE database. Gathered biological information is required to run the 3 other modules of the workflow: "Gene Ontology", "Secreted Proteins", and "Protein Interaction" (described in the "Results" section of this article). The

GO program queries QuickGO and ProteINSIDE's databases to perform the functional annotation. The GO program analyses over- and under-represented terms to highlight the most relevant GO terms related to the input. These statistical calculations are made with an R script performing a Fisher's exact test (functional enrichment first proposed by FatiGO [30]) and the resulting *p*-value is corrected or not by the Benjamini & Hochberg (BH) test [31]. The prediction of secreted proteins is made using a local version of SignalP (version 4.1) that looks for a signal peptide on amino acid sequence of each protein [16] (cutoff of 0.45 and 0.34 for SignalP prediction, in the basic and custom analysis with the sensitive option selected, respectively; for more information see the tutorials of SignalP^a). To ascertain that proteins are secreted, ProteINSIDE uses TargetP [22] (version 1.1) to predict the cellular location of each protein. ProteINSIDE uses a pre-set cutoff option to get a significant prediction (higher than 95%) according to TargetP instructions^b. Protein interactions are searched using PSICQUIC service [17] and statistical calculations are made with an R script and the package "tnet" [32]. ProteINSIDE performs sequence alignment thanks to a local version of NCBI BlastP [33] against UniProt/Swissprot databases [24, 34]. Lastly and as an additional valuable tool, ProteINSIDE lists in one table all known IDs for an input of proteins or genes thanks to the ProteCONVERT module. This list is the result of a search and of a gathering of IDs thank to the ProteINSIDE biological database. Only registered users have access to the ProteCONVERT module.

Figure 2. Setting up a basic analysis. First, enter a name for the analysis and select the species of study. There are two ways to submit a protein or gene list; you can use an input file or directly paste your IDs. The input file must be less than 250 kb and the file format must be specified. There is also a "Sample" button that loads parameters for an example analysis. Once everything is filled, click on the button "Run the job" to submit.

The results of an analysis with ProteINSIDE are available online or downloadable. Unregistered users get results through a unique code and a link provided after the submission of a list. Registered users have access to the results through their analysis manager after their connection with their login and their password. Four separated pages provide results from the four analyses ("ID Mapping", "Gene Ontology", "Secreted Proteins", and "Protein Interaction"). The results are dynamic tables and charts that can be sorted and filtered by specific criterion such as biological function, protein or gene ID, or *p*-value. Tables and charts are made using Google Charts package (<https://developers.google.com/chart/>) and are downloadable, diagrams and histograms are printable. Networks are downloadable as image (.pdf or .png) or as network viewer input files (Cytoscape .cys or graphml and xgmml). The results interpretations are detailed in the results section.

Implementation

ProteINSIDE is freely available online at www.proteinside.org and doesn't require an installation on a computer. ProteINSIDE is completely adapted for any internet browser and tablet. We recommend multiprocessors computer with at least 2 GB of ram to get better performances for huge network visualization and filtering.

The web interface is programmed in PHP, HTML, and JavaScript. The workflow has been completely programmed in Perl (version 5.10.1; CPAN modules (Comprehensive Perl Archive Network) and BioPerl [35]) and R scripts (version 3.0.1). The database was made in MySQL (version 5.5) (Figure 1).

Sample dataset

We have created a dataset to assess ProteINSIDE performances. This dataset is composed of the UniProt accession numbers of 133 proteins (Table 1): 34 proteins related to the glycolysis cycle, 11 proteins from the respiratory chain, 5 proteins from the tricarboxylic acid cycle (TCA), 79 hormones or secreted proteins, and proteins with very specific functions unrelated to the others. We also included a duplicated ID among proteins of the glycolysis to verify its recognition by ProteINSIDE. ProteINSIDE is able to detect duplicate protein even if the IDs are different: a Gene Name, a UniProt accession number, and a Gene Identifier related to a same protein will be taken into account as a single protein.

We have created this dataset on bovine species, but the number of annotations and PPi weren't sufficient for a complete overview of the functionalities of ProteINSIDE. Then, we used the same proteins using human IDs to test ProteINSIDE with the basic and the custom analyses (Table 1).

RESULTS

Here we present how to run a basic or custom analysis and how to view the results. We explain how to interpret the results and we discuss the relevance of biological information extracted by ProteINSIDE for our sample dataset of 133 proteins.

Setting up a Basic Analysis: a standard overview of a dataset

ProteINSIDE performs a basic analysis (in which settings are locked and the workflow provides GO terms, list of putative secreted proteins, and PPi data from IntAct [36], UniProt [24], and BioGrid [37] databases). A basic analysis gives a complete overview of a dataset. To set up a basic analysis, user has to follow these steps (Figure 2):

- Click on "Basic Analysis" menu on the homepage of ProteINSIDE
- Fill in "the job name" box
- Select the species of study (the same species as the uploaded IDs)
- Upload an input file or directly paste IDs
- Click on the "Run the job" button to submit a new analysis

The analysis status is indicated by the colour of a button: red for "analysis on the waiting list", yellow for "the analysis is running" and green "analysis done". The blue globe is the link to access to the online results:

- Click on the blue globe button to view the results (or use the trash to delete them)
- Visualise the results summary produced by the four modules of analysis on the first default page (entitled "Results Summary", Figure 3)
- Navigate between the four module's results pages by clicking on the module's name on the toolbar menu.

Analyses and data	Glycolysis	Hormones	TCA	Analysis duration (min)
Dataset	33+1 duplicated	79	5	-
Basic analysis	27	78	3	2
Custom analysis	33	79	5	10

Table 1. Results summary of ProteINSIDE analysis performances. The numbers are the proteins that belong to main pathways in the sample dataset, that are properly annotated by GO terms relevant to glycolysis and tricarboxylic acid (TCA) pathways, and that have been predicted as secreted by SignalP (and confirmed by GO terms, TargetP, and subcellular location) for hormones.

For our sample dataset, the "Results Summary" page reported that all 133 proteins were recognized by ProteINSIDE and the protein in duplicate was identified and excluded from the analysis (Figure 3). Thus, 132 proteins were submitted to the four modules of analysis.

The "ID Mapping's" module aimed to retrieve and gather basic biological knowledge, results are directly viewed on the "ID Resume" web page of ProteINSIDE. This module compares each submitted IDs to the database of ProteINSIDE to ascertain a match with genes or proteins from human, rat, mouse, cow, sheep or goat species. The local biological database of ProteINSIDE is a combination of NCBI Gene/Protein, NCBI HomoloGene [28], and UniProt [24] databases. These databases were chosen because data are easily extractable, curated and daily updated. For each uploaded ID, ProteINSIDE obtains and summarises as a downloadable table (Figure 4): gene or protein ID, gene or protein name, a summary of protein function, gene chromosomal location, and information on tissue expression and cellular location. The module also recovers the protein sequence of each input ID. Each protein and gene ID listed on this web page are linked to corresponding UniProt and NCBI web pages. FASTA amino acid sequences of each input are also downloadable.

The module dedicated to the functional annotation according to the GO consortium, produces results that are viewed on the "GO" web page of ProteINSIDE. ProteINSIDE imports GO terms by querying the QuickGO database [38]. QuickGO was chosen because of its daily update, accessibility, and performances. ProteINSIDE only imports GO terms that have been selected by evidence codes (GO Inferred from Electronic Annotation codes (IEA) are excluded by the basic analysis) and confirmed by

curators. The GO script of ProteINSIDE analyses over- and under-represented terms to identify the most relevant and the most specific terms associated with the uploaded list. Within a GO, ProteINSIDE compares the number of genes or proteins from the dataset to the total number of gene products (for a species) declared in the AmiGO database [29] to provide a coverage frequency, and thus, to identify the most representative pathways associated to a dataset. The result is viewed on the "GO" web page of ProteINSIDE as tables and diagrams. Three tables (Figure 5) report the GO terms that annotated two or more proteins (Figure 5-B), the GO terms that annotate one protein (Figure 6-C), as well as all GO terms for a protein (Figure 5-D). Each annotation is informed with an evidence code (that reflects the type of experimental evidence or analysis to describe an annotation between a GO term and a gene product) and the database source. Tables are automatically sorted by the best enrichment *p*-value to help the user to view the most significant GO terms related to a dataset. Tables can also be sorted by ontology group, *p*-value range for enrichment, GO term description, gene name or any input IDs (Figure 5B). From the sample dataset of 132 proteins, ProteINSIDE annotated 128 proteins with 624 GO terms. The most significant enriched GO terms is "hormone activity" (that annotated 31 proteins over the 79 expected; not shown) and "glycolytic process" (that annotated 27 proteins over the 33 expected; Table 1). The low number of annotated proteins may be related to our choice to use only GO terms that have been confirmed by curator in the basic analysis. This means that the basic analysis doesn't use the annotation with IEA (Inferred by Electronic Annotation) evidence code. However, the option to use IEA is provided in the custom analysis to extend the annotations.

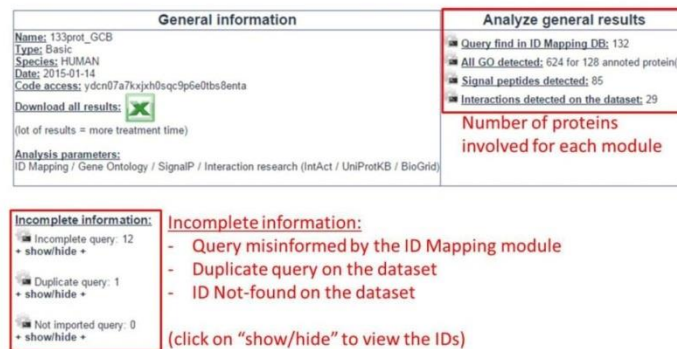


Figure 3. Main page of results produced by a basic analysis. This is the first page of the results. It shows the number of proteins or genes successfully analysed by each module.

The module that aims to predict potentially secreted proteins provides results on the “Secreted Proteins” web page of ProteINSIDE (Figure 6). To identify proteins that are putatively secreted, ProteINSIDE first predicts the presence of a signal peptide on a protein sequence (imported by the “ID Mapping” module) through a local version of the SignalP tool [16]. SignalP was chosen because of its high prediction score in comparison with other available tools [39, 40]. To ascertain the prediction, a local version of TargetP software [22] predicts the subcellular location of the proteins. ProteINSIDE also checks the subcellular location of the protein using UniProt source to confirm TargetP results. As a final verification step, ProteINSIDE selects the GO terms related to secretory pathways for each SignalP prediction. For this purpose, we have selected about 1,000 GO terms related to secretion (monthly updated) as for example: secretion, vesicle, or extracellular region. This four-step analysis improves the reliability of proteins proposed to be secreted thanks to a signal peptide and to our knowledge is unique to ProteINSIDE [40]. However, proteins are also secreted by pathways that do not involve signal peptide such as: endosomal recycling, plasma membrane transporter, membrane flip-flop, and membrane blebbing including the formation of vesicles or exosomes [41]. Thus, ProteINSIDE was designed to predict the proteins secreted by other pathways, by gathering the data of subcellular location provided by UniProt, GO terms, and TargetP results (Figure 6-B). From our sample dataset of 132 proteins, ProteINSIDE has predicted 85 proteins as potentially secreted outside the cell by a signal peptide, among them 78 over the 79 proteins that were expected (Table 1). This lack of perfect prediction can be explained by the false positive and false negative prediction rates of SignalP, as already evaluated by Petersen et al. (Supplementary materials and methods of [16]). Over the 85 predicted secreted proteins, 65

were also annotated by GO terms related to the secretion. The subcellular locations of 81 proteins were both confirmed by TargetP and UniProt source. Additionally, 30 proteins were predicted to be secreted without signal peptide.

The fourth module is dedicated to PPI analysis and results are viewed on “Protein Interaction” web page of ProteINSIDE. PPI identification and visualisation within a network conveyed how various genes or proteins contribute to cellular or metabolic processes. ProteINSIDE uses the PSICQUIC service [17] to identify PPI and imports PPI identified by their “interaction detection methods” with experimental proofs and confirmed by curator. The basic analysis identifies PPI within the uploaded dataset (core network) using the preselected databases IntAct, UniProt, and BioGrid. These PPI databases were chosen as a default option because there are daily updated and reviewed by curators as well as by the curation processes of the IMEx project (that ensures reliable interactions data using experts and curation rules shared between many interaction databases [26]) or MIMIX (a guideline of the minimum information required for reporting a molecular interaction experiment, thus advising the user on how to use the interaction data [27]). Moreover, BioGrid is the biggest PPI database that has its own curation workflow (more than 740000 curated PPI) and is not a partner of IMEx curation program. IntAct is another big PPI database with more than 380000 PPI curated according to IMEx and MIMIX curation rules and that are often listed in several databases. UniProt is a major database dedicated to the study of proteins. Thus, it possesses its own curated PPI but in lesser amounts compared to the two other specialized databases (less than 13 000 PPI; UniProt is a partner of IMEx project). By using 3 databases as a default option, the aim of ProteINSIDE is to favour the use of multiple PPI databases in order to improve the PPI data gathering [42].

Research area						Download results		
Gene Name	Protein	Function	Tissue	Subcellular location	Number of rows	download table	Sequences (FASTA)	
Query	Proteins	Protein ID	Gene Name	Gene Entrez	Function	Chromosome	Tissue	Subcellular location
1	Q1568	ADIPO_HUMAN	ADIPOG	9275	Important adipokine involved in the control of fat metabolism and insulin sensitivity, with direct anti-diabetic, anti-atherogenic and anti-inflammatory activities. Stimulates AMPK phosphorylation and activation in the liver and the skeletal muscle, enhancing glucose utilization and fatty-acid combustion. Antagonizes TNF- α by negatively regulating its expression in various tissues, such as liver and macrophages, and also by counteracting its effects. Inhibits endothelial NF- κ B signaling through a cAMP-dependent pathway. May play a role in cell growth, angiogenesis, and tissue remodeling by binding and sequestering various growth factors with distinct binding affinities, depending on the type of complex. LIGY, MBOY, or HMOY.	3q27 NC_000003.12 (19642012..19665463)	Synthesized exclusively by adipocytes and secreted into plasma	Secreted
2	Q724H4	ADM2_HUMAN	ADM2	7992	ADM1 and ADM2 may play a role as physiological regulators of gastrointestinal, cardiovascular bioactivities mediated by the CALCRL/RAMP3 receptor complex. Activates the cAMP-dependent pathway.	22q13.33 NC_000022.11 (58481656..58486437)	Expressed in the esophagus, stomach, jejunum, ileum, ileocecum, ascending colon, transverse colon, descending colon and rectum. Expressed in myocardial cells of the heart, renal tubular cells, hypothalamus, and pituitary.	Secreted
3	P35310	ADM1_HUMAN	ADM	133	ADM and PAMP are potent hypertensive and vasodilator agents. Numerous actions have been reported most related to the physiologic control of fluid and electrolyte homeostasis. In the kidney, ADM is diuretic and natriuretic, and both ADM and PAMP inhibit aldosterone secretion by direct adrenergic actions. In pituitary gland, both peptides at physiologically relevant doses inhibit basal ACTH secretion. Both peptides appear to act in brain and pituitary gland to facilitate the loss of plasma volume, actions which complement their hypertensive effects in blood vessels.	11p15.4 NC_000011.10 (10395095..10397276)	Highest levels found in atrioventricular and adrenal medulla. Also found in lung, ventricle and kidney tissues.	Secreted

Dynamic table: results are sorted and filtered by clicking on columns; a specific search is made possible by a search depending on gene name, protein accession number or protein function.

Figure 4. Biological knowledge retrieval. The ID Mapping module results are listed in a table. This table provides protein IDs, gene names, summaries the protein function, chromosomal locations, data on tissue expression, and subcellular location. User can sort the table by using the dynamic table research area.

These 3 PPI databases ensure the good recovery of known interactions for an overview of interactions within or/and outside of a new dataset. Then, ProteINSIDE lists pairs of proteins known to interact between each other in a downloadable table (Figure 7) and constructs a network (Figure 8) using the PPI identified within the uploaded list. The dynamic network is available by using the "Cytoscape" button on the "Protein Interaction" web page ("Dynamic Cytoscape view of PPI", Figure 7-A). Within the network, edges are experimental detection methods used to identify the PPI. Consequently, several edges may link two proteins. Network can be sorted by the number of interactions by node, the proximity of a node to other nodes (closeness centrality; CC) and the shortest paths between nodes (betweenness centrality; BC) (Figure 8-A). These centralities criteria were already proven to be efficient to select key nodes/proteins within a pathway [43]. From our sample dataset of 132 proteins, ProteINSIDE has identified 29 PPI that involved 28 different proteins (Figure 7-B). As expected from our small dataset, ProteINSIDE linked, within sub networks, proteins involved in glycolysis, TCA or respiratory chain as protein complexes (partially on Figure 8-B).

Setting up a Custom Analysis: an added-value provided by the extension of the analysis

We made a custom analysis using the same major settings as for the basic analysis with additional options (GO network, GO electronic annotations, and extension of PPI to proteins outside of the dataset in the same species, extended network). To set up the custom analysis, user has to follow these steps (also explained by Figure 2):

- Click on the "Custom Analysis" menu on the homepage of ProteINSIDE
- Fill in "the job name" box
- Select the species of study (the same species as the uploaded IDs)
- Upload an input file or directly paste IDs

Then, user has to select the settings of either all or only one module of analysis on the section "4" of the page, by following these steps (Figure 9):

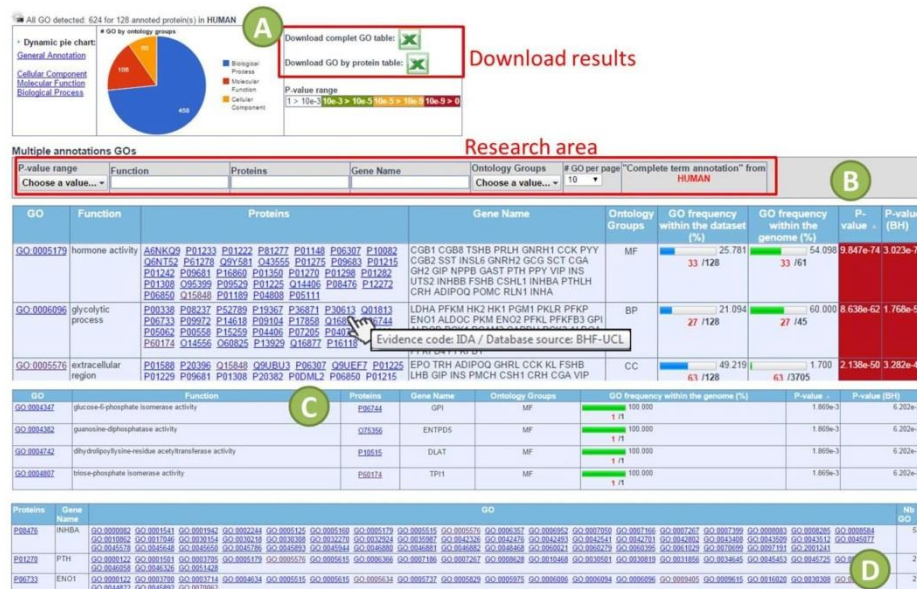


Figure 5. Functional annotation according to the Gene Ontology. GO results are first extracted and classified by the number of GO terms related to Molecular Functions, Biological Processes, and Cellular Components, then visualised as diagrams or downloadable tables. (A) Main menu of GO results page, to download the results as Excel files, to view the significance of p-value range colours, or a proportion of major annotation categories as diagram. (B and C) GO terms are also sorted as two dynamic tables (a table for GO terms that annotate more than one protein on the dataset - B, and a second table for the GO terms with a single protein annotated - C). Tables can be sorted by GO term, function, protein name or ID, gene name, number of annotations, annotation frequency or annotation enrichment. (D) A third table lists all GO codes for a given protein. Users can move the cursor over a protein to be informed about the evidence code and the database source of the GO annotation (B; where IDA means "Inferred from Direct Assay").

- Select Gene Ontology module
 - Select "GO electronic annotation (IEA)" if you aim to use GO annotation inferred from electronic annotation.
 - Select "Gene Ontology Tree network" to view linked GO terms
- Select Signal peptide module to use the basic cutoff value of prediction
- Select Protein - protein interaction module
 - Select "Protein - protein interaction custom analysis"
 - Select "Extend PPI research using protein outside of the dataset", if wanted
 - Select "Human species" to analyse PPI using data available in Human, for example
 - Select either the 3 most used databases (IntAct, UniProt, and BioGrid as used in the basic analysis) or from 1 to 31 databases (PPI are daily updated in each database)

Alternatively, user can load automatically the same settings as those already used in a previous custom analysis, by clicking on the "pre-set" button.

At the completion of the custom analysis, the "ID Resume" web page provides the same information than the basic analysis (Figure 3).

Within the GO module, the choice to use the electronic annotation option has increased both the number of annotated proteins (132 rather than 128 without IEA in the basic analysis) and the number of annotations by around 40% (1080 unique GO terms rather than 624 in the basic analysis). Thanks to IEA option, ProteINSIDE correctly retrieved the 33 expected proteins related the glycolytic process and the 79 proteins related to a hormone activity (Table 1). The GOTree network linked 570 GO terms. A link between 2 terms is represented by an "is_a" relation: "Diuretic hormone activity" linked to "Hormone activity" means that "Diuretic hormone activity" is a "Hormone activity" pathway. The network can be sorted by ontology group, by *p*-value range (to select and to link only the most enriched GO terms), by the number of directly linked terms or also by the number of GO terms linked together (to select group of GO terms involved in the same biological function). From our sample dataset, we have chosen to illustrate the GO tree of the "Molecular Function" group (Figure 10). In this visualisation, squares with dark red colour were GO terms which have annotated the highest number of proteins. Among them and as expected the GO:0005179 with the best *p*-value and the darkest red colour was "Hormone activity", in agreement with the over representation of hormones in our sample dataset.

The "Secreted Proteins" module has predicted the same 85 proteins as the basic analysis as being secreted. By comparison with the basic analysis, the use of IEA option has allowed to confirm this prediction for 82 proteins that were also annotated with GO terms related to a "secretion" function.

By comparison with the basic analysis, the settings selected within the "Protein Interaction" module provided PPI within the dataset (between proteins of the dataset, as the basic analysis) and PPI between proteins from the dataset and outside of the dataset. For the extended network, ProteINSIDE retrieved 688

PPI made by 500 proteins. Among them, 61 proteins were from our uploaded sample dataset. By using PPI outside of the dataset in Human species, we got 95% more PPI that involved 60% more proteins from the sample dataset than the PPI recovered with the basic analysis. The extended network (Figure 11) highlighted major subnetworks related to the respiratory chain (Figure 11-A), hormone activity such as signalization pathways of adipokines (Figure 11-B), growth hormone (Figure 11-C), thyroid hormones (Figure 11-D), glycolysis (not highlighted), and carbohydrate metabolism (not highlighted). This is consistent with the over selection of proteins from glycolysis, TCA or hormones or adipokines. Betweenness and closeness centralities were used to sort the most central proteins of this extended network (Figure 11-E). By this way, we identified 22 highly central proteins, 13 of them coming from the uploaded sample dataset and involved in respiratory chain and glycolysis as protein complexes.

DISCUSSION

Currently, most genomic and proteomic studies increasingly generate data which have to be gathered, filtered, and analysed using one or more softwares [44-46]. The major and widely used strategies to systematically study proteins [47] and genes [48] in a cell are based on functional annotation, proteins interactions and pathways analysis. The literature describes many tools for genomic and proteomic data analysis [4]. Scientists have to select appropriate tools among those for either the GO annotation [15, 30, 49, 21, 29, 50], the prediction of secreted proteins [51, 52, 39, 53], or the search of protein - protein interactions [54, 55, 36, 56, 37, 57, 58].

ProteINSIDE is not just an additional resource since it was designed to provide efficient and original strategies to run in a single query, biological knowledge gathering, GO terms annotation, secreted protein prediction, and protein interaction. The DAVID [59], ToppGene [12] or Babelomics [60] software resources are often mentioned for the biological knowledge gathering, functional annotation using GO terms or searches for proteins interactions. By comparison to these tools, added-values of ProteINSIDE have to be highlighted.

ProteINSIDE provides a functional annotation using a monthly updated GO terms database and enrichment calculation. Indeed, the list of GO terms is in constant evolution and GO terms could become redundant or obsolete the next month [15]. This could induce bad information in the results of an analysis if the database is not often updated. Each result of the annotation is easily readable thanks to dynamic tables and diagrams which can be sorted with many options and can be downloaded to work offline. The GO tree visualization of the most often associated GO terms with a list of IDs, is another added-value of ProteINSIDE. Tree networks of GO terms are also done by AmiGO or QuickGO to get an ancestor chart of a single term. However, ProteINSIDE is the only tool which highlights biological pathways of a dataset using linked GO terms and their representativeness rate (using *p*-values and number of annotations). This network visualization is also easy to use thanks to the friendly user interface that gives access to the sort

options. For the PPI research and visualization, ProteINSIDE uses only interactions that are based on experimental observations. The drawback is that the number of PPI identified by ProteINSIDE could be lower than those proposed by other resources that also list predicted interactions inferred from literature mining. Furthermore, ProteINSIDE is also capable to draw large interaction networks thanks to the use of the powerful graphical Cytoscape application. ProteINSIDE provides different options to filter large networks, making it as easy to use as the widely used resource STRING [57], and efficient to select keys proteins in a network. Moreover, to analyse locally the networks, files (e.g. .cys, xgmml, graphml) are ready to be open by a network viewer like Cytoscape (and its numerous plugins) and are downloadable from the PPI page result. To our knowledge, among the tools to mine genomics data from mammals, ProteINSIDE is the only resource that allows a very simple view and analysis of network, and prepares data for their further download and analysis by other network viewer software as Cytoscape. These features may be valuable for biologists without a strong bioinformatics background. For the less informed species, ProteINSIDE allows searching PPI in well-informed species thanks to homologous IDs. For this, ProteINSIDE automatically selects homologous IDs from its database for the wanted species. Nevertheless, user can choose to run a local Blastp to select the species with the highest sequence homology with the proteins of the input dataset, and then ProteINSIDE proceed to the selection of orthologous IDs for this species. A functional annotation of all proteins from an extended network (PPI between proteins within and outside of the dataset) is done by clicking a button on the network visualisation. Results of this annotation are available as a new analysis. In addition to biological knowledge gathering, GO annotation, and analysis of PPI, ProteINSIDE also proceeds to an *in silico* secretome analysis [40]. For this purpose, ProteINSIDE merges four strategies of analysis: signal peptide [16] and cellular location [22] predictions, as well as a review of GO term annotation and cellular location recorded in UniProt. This four-step analysis provides a reliable prediction of proteins secreted thanks to a signal peptide. To our knowledge, ProteINSIDE is the unique all-in-one tool that predicts secretome from a list of gene or protein IDs [40].

Scientists are dependent on the species of study when they choose among resources available for their genomic and proteomic data analysis. Indeed, many tools are dedicated to only one species such as BioMyn for the Human [9] or DroPNet for the *Drosophila* [7]). Moreover, many tools are dedicated to diseases studies such as NetPath [13] and ToppGene [12]. ProteINSIDE has been first tool designed for genomic and proteomic data analysis in ruminant species namely cattle, sheep, and goat. However, the lack of information on these species required us to add human, rat, and mouse species to do homologous analysis. Thus, IDs from these species are perfectly recognized and analysed by ProteINSIDE. To our knowledge, ProteINSIDE is the only resource that allows the user to recover biological knowledge from well-known species (human, rat or mouse) using IDs from ruminant species. This avoids losing information

since many sequences or annotations remain to be stored in public databases for ruminant species and especially for goat. To our knowledge, only AgBase [61], a manually curated gene annotation database for farm species, including cattle and sheep, is available for functional annotation. However, AgBase does not perform analysis of PPI or prediction of secreted proteins.

In this article we have presented the performances of ProteINSIDE, a new powerful workflow which gathers tools and public databases to retrieve biological information of genes or proteins lists from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). We have reported a tutorial to describe how to get and interpret the results of a basic and a custom analysis with ProteINSIDE. Currently, there is no tool that performs in one query the analyses proposed by ProteINSIDE. ProteINSIDE offers a friendly-user interface where user can view, work, and download the results of an analysis. ProteINSIDE gives also a single file containing all results of an analysis. Thus, ProteINSIDE offers a great support to analyse efficiently a large quantity of data from genomic and proteomic studies to gather and interpret results necessary to construct a new research hypothesis or answer to a single question.

ACKNOWLEDGEMENTS

N. Kaspric's PhD grant is provided by the regional council of Auvergne in France, APIS-GENE and the regional information system Lifegrid with the help of Feder has provided the grant of J. Tournayre.

AUTHOR CONTRIBUTIONS

Conceive and design ProteINSIDE web service: NK, JT, MR, and MB. Perform experiments and analysed the data: NK and MB. Write the paper: NK, MB, and BP.

CONFLICT OF INTEREST DECLARATION

The authors declare no conflict of interest.

SUPPLEMENTARY DATA

High resolution files of the main figures on the paper are available for download at [Genomics and Computational Biology online](#).

ABBREVIATIONS

BC: betweenness centrality
 CC: closeness centrality
 GO: gene ontology
 ID: identifier
 PPI: protein-protein interaction

options. For the PPI research and visualization, ProteINSIDE uses only interactions that are based on experimental observations. The drawback is that the number of PPI identified by ProteINSIDE could be lower than those proposed by other resources that also list predicted interactions inferred from literature mining. Furthermore, ProteINSIDE is also capable to draw large interaction networks thanks to the use of the powerful graphical Cytoscape application. ProteINSIDE provides different options to filter large networks, making it as easy to use as the widely used resource STRING [57], and efficient to select keys proteins in a network. Moreover, to analyse locally the networks, files (e.g. .cys, xgmml, graphml) are ready to be open by a network viewer like Cytoscape (and its numerous plugins) and are downloadable from the PPI page result. To our knowledge, among the tools to mine genomics data from mammals, ProteINSIDE is the only resource that allows a very simple view and analysis of network, and prepares data for their further download and analysis by other network viewer software as Cytoscape. These features may be valuable for biologists without a strong bioinformatics background. For the less informed species, ProteINSIDE allows searching PPI in well-informed species thanks to homologous IDs. For this, ProteINSIDE automatically selects homologous IDs from its database for the wanted species. Nevertheless, user can choose to run a local Blastp to select the species with the highest sequence homology with the proteins of the input dataset, and then ProteINSIDE proceed to the selection of orthologous IDs for this species. A functional annotation of all proteins from an extended network (PPI between proteins within and outside of the dataset) is done by clicking a button on the network visualisation. Results of this annotation are available as a new analysis. In addition to biological knowledge gathering, GO annotation, and analysis of PPI, ProteINSIDE also proceeds to an *in silico* secretome analysis [40]. For this purpose, ProteINSIDE merges four strategies of analysis: signal peptide [16] and cellular location [22] predictions, as well as a review of GO term annotation and cellular location recorded in UniProt. This four-step analysis provides a reliable prediction of proteins secreted thanks to a signal peptide. To our knowledge, ProteINSIDE is the unique all-in-one tool that predicts secretome from a list of gene or protein IDs [40].

Scientists are dependent on the species of study when they choose among resources available for their genomic and proteomic data analysis. Indeed, many tools are dedicated to only one species such as BioMyn for the Human [9] or DroPNet for the *Drosophila* [7]). Moreover, many tools are dedicated to diseases studies such as NetPath [13] and ToppGene [12]. ProteINSIDE has been first tool designed for genomic and proteomic data analysis in ruminant species namely cattle, sheep, and goat. However, the lack of information on these species required us to add human, rat, and mouse species to do homologous analysis. Thus, IDs from these species are perfectly recognized and analysed by ProteINSIDE. To our knowledge, ProteINSIDE is the only resource that allows the user to recover biological knowledge from well-known species (human, rat or mouse) using IDs from ruminant species. This avoids losing information

since many sequences or annotations remain to be stored in public databases for ruminant species and especially for goat. To our knowledge, only AgBase [61], a manually curated gene annotation database for farm species, including cattle and sheep, is available for functional annotation. However, AgBase does not perform analysis of PPI or prediction of secreted proteins.

In this article we have presented the performances of ProteINSIDE, a new powerful workflow which gathers tools and public databases to retrieve biological information of genes or proteins lists from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). We have reported a tutorial to describe how to get and interpret the results of a basic and a custom analysis with ProteINSIDE. Currently, there is no tool that performs in one query the analyses proposed by ProteINSIDE. ProteINSIDE offers a friendly-user interface where user can view, work, and download the results of an analysis. ProteINSIDE gives also a single file containing all results of an analysis. Thus, ProteINSIDE offers a great support to analyse efficiently a large quantity of data from genomic and proteomic studies to gather and interpret results necessary to construct a new research hypothesis or answer to a single question.

ACKNOWLEDGEMENTS

N. Kaspric's PhD grant is provided by the regional council of Auvergne in France, APIS-GENE and the regional information system Lifegrid with the help of Feder has provided the grant of J. Tournayre.

AUTHOR CONTRIBUTIONS

Conceive and design ProteINSIDE web service: NK, JT, MR, and MB. Perform experiments and analysed the data: NK and MB. Write the paper: NK, MB, and BP.

CONFLICT OF INTEREST DECLARATION

The authors declare no conflict of interest.

SUPPLEMENTARY DATA

High resolution files of the main figures on the paper are available for download at [Genomics and Computational Biology online](#).

ABBREVIATIONS

BC: betweenness centrality
 CC: closeness centrality
 GO: gene ontology
 ID: identifier
 PPI: protein-protein interaction

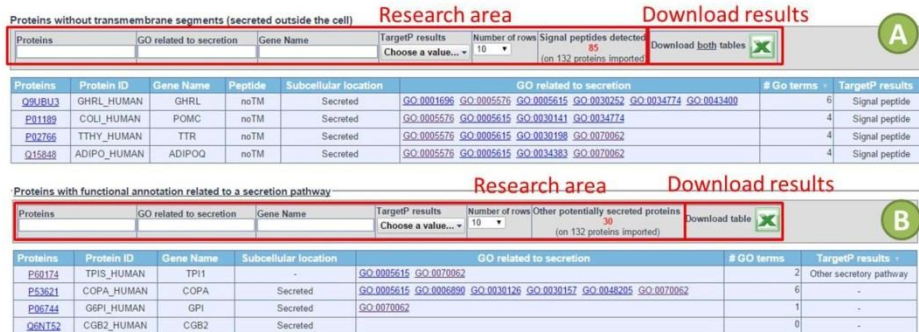


Figure 6. Prediction of secreted proteins. Proteins potentially secreted are listed as two or three downloadable dynamics tables. (A) The first table lists proteins predicted as secreted by SignalP. The column "Peptide" provides the results for a positive identification of a signal peptide on a protein sequence as provided by SignalP. Identified peptides can be "noTM" (not transmembrane) or "TM" (transmembrane), only "noTM" are listed in the first table. The column "Subcellular location" provides the location of the protein declared in the UniProt database. The column "TargetP" provides the prediction of the subcellular location of the protein by TargetP software, and GO related to secretion are also listed to improve the prediction. A second table lists proteins with the "TM" prediction of SignalP, not shown in the figure since there was no result with the sample dataset. (B) A third table lists proteins potentially secreted by secretory pathways that do not involve signal peptide. In this table, GO terms, TargetP prediction, and subcellular location are also selected to improve the prediction.

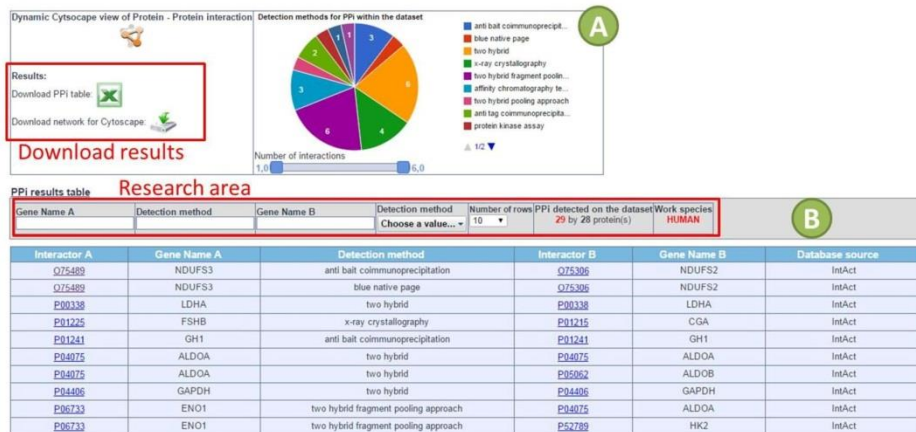


Figure 7. PPI results and visualisation. Results for PPI are summarised as a downloadable table and a diagram. (A) Main results are downloadable as table and network file that can be visualized using a network viewer (as Cytoscape). An online network view (made using the Cytoscape web application) is also proposed from this page result. A pie diagram indicates the number of PPI identified with the different detection methods. (B) A dynamic table lists linked proteins within the dataset, the detection method used to identify the interaction, and the database source of the interaction.

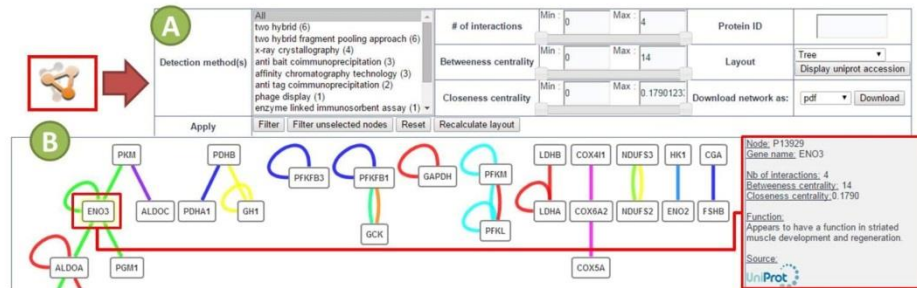


Figure 8. Network visualization of the PPI results. (A) This menu provides options to filter the network by: detection method, number of interactions for a protein, type of layout, protein ID, or the values of centralities. The centralities values are useful to sort large networks and to view only a central subnetwork. The betweenness centrality quantifies how frequently a node is on the shortest path between every pair of nodes for detecting bottlenecks in a network. The closeness centrality quantifies how distant minimal paths are from a given node to all others, a large closeness indicates that a node is close to the topological center of the network. (B) The network view is a dynamic image where user can access to a protein data by clicking on a node (name, function, statistic results, and database source and link of the protein).



Figure 9. Setting up a custom analysis. Firstly, user has to enter a name for the analysis, select the species of study, and directly paste the input IDs (Figure 2). User has to select settings of the analysis: the setting followed by "software" mention activates the corresponding module in the workflow, and then user can select options for chosen module(s).

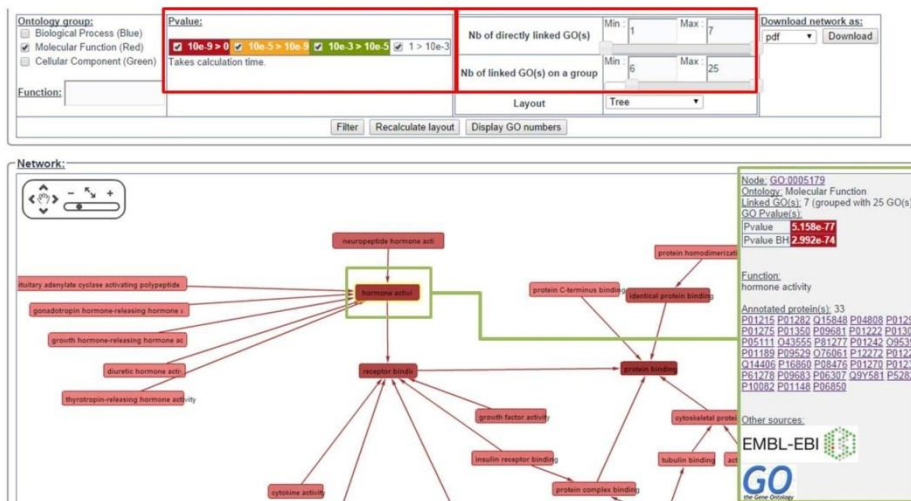


Figure 10. GOTree network visualization. Linked GO terms which annotate the dataset are linked using ancestor chart method. Each edge means that a term A is a subtype of a term B (is_a). Information about a GO is obtained by clicking on the GO or the node. Red colour is only for the GO terms relative to the Molecular Function. 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).

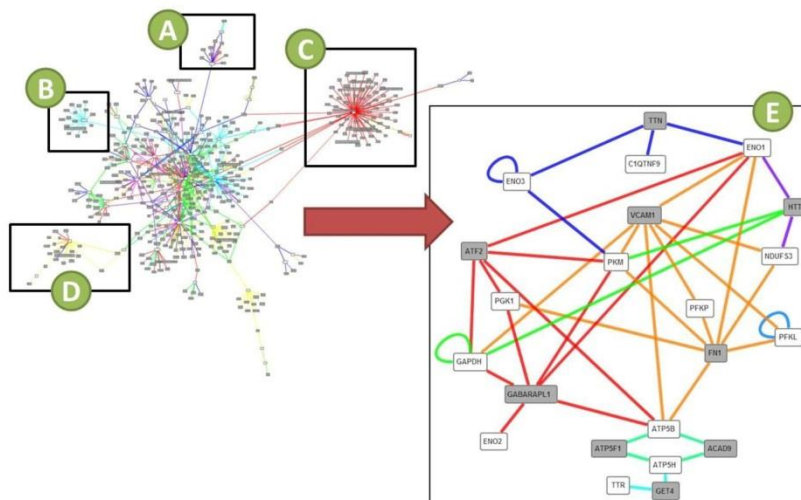


Figure 11. Extended network of PPI with proteins outside of the dataset. This network is made of PPI retrieved by querying the BioGrid, UniProt, and IntAct databases and using PPI with human proteins outside of the dataset. Grey squares are for proteins outside the dataset; white proteins are from the dataset. We have highlighted linked proteins that are involved in pathways such as: (A) glycolysis, (B) hormone activity, (C) the growth hormone signalling, and (D) thyroid hormones signalling. (E) We have used high values of betweenness and closeness centralities (BC: 3600; CC: 0.2) to get the most central proteins of this extended network.

REFERENCES

- Chaze T, Meunier B, Chambon C, Jurie C, Picard B. **Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle**. *Animal: an international journal of animal bioscience*. 2009;3(7):980-1000. doi:[10.1017/S1751731109004315](https://doi.org/10.1017/S1751731109004315).
- Picard B, Cassar-Malek I, Guillemain N, Bonnet M. **Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production**. In: Moo-Young M, editor. *Comprehensive Biotechnology (Second Edition)*. Burlington: Academic Press; 2011. p. 395-405.
- Taga H, Chilliard Y, Meunier B, Chambon C, Picard B, Zingaretti MC et al. **Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown?** *Journal of cellular physiology*. 2012;227(4):1688-700. doi:[10.1002/jcp.22893](https://doi.org/10.1002/jcp.22893).
- Schmidt A, Fome I, Imhof A. **Bioinformatic analysis of proteomics data**. *BMC Syst Biol*. 2014;8 Suppl 2:S3. doi:[10.1186/1752-0509-8-S2-S3](https://doi.org/10.1186/1752-0509-8-S2-S3).
- Wong AK, Park CY, Greene CS, Bongo LA, Guan Y, Troyanskaya OG. **IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks**. *Nucleic Acids Res*. 2012;40(Web Server issue):W484-90. doi:[10.1093/nar/gks458](https://doi.org/10.1093/nar/gks458).
- Pache RA, Ceol A, Aloy P. **NetAligner—a network alignment server to compare complexes, pathways and whole interactomes**. *Nucleic Acids Res*. 2012;40(Web Server issue):W157-61. doi:[10.1093/nar/gks446](https://doi.org/10.1093/nar/gks446).
- Renaud Y, Baillif A, Perez JB, Agier M, Mephu Nguifo E, Mirouse V. **DroPNet: a web portal for integrated analysis of Drosophila protein-protein interaction networks**. *Nucleic Acids Res*. 2012;40(Web Server issue):W134-9. doi:[10.1093/nar/gks434](https://doi.org/10.1093/nar/gks434).
- Tuncbag N, McCallum S, Huang SS, Fraenkel E. **SteinerNet: a web server for integrating 'omic' data to discover hidden components of response pathways**. *Nucleic Acids Res*. 2012;40(Web Server issue):W505-9. doi:[10.1093/nar/gks445](https://doi.org/10.1093/nar/gks445).
- Ramirez F, Lawyer G, Albrecht M. **Novel search method for the discovery of functional relationships**. *Bioinformatics*. 2012;28(2):269-76. doi:[10.1093/bioinformatics/btr631](https://doi.org/10.1093/bioinformatics/btr631).
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R et al. **The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools**. *Nucleic Acids Res*. 2012;40(Database issue):D1202-10. doi:[10.1093/nar/gkr1090](https://doi.org/10.1093/nar/gkr1090).
- Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C et al. **EcoCyc: fusing model organism databases with systems biology**. *Nucleic Acids Res*. 2013;41(Database issue):D605-12. doi:[10.1093/nar/gks1027](https://doi.org/10.1093/nar/gks1027).
- Chen J, Bardes EE, Aronow BJ, Jegga AG. **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization**. *Nucleic Acids Res*. 2009;37(Web Server issue):W305-11. doi:[10.1093/nar/gkp427](https://doi.org/10.1093/nar/gkp427).
- Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK et al. **NetPath: a public resource of curated signal transduction pathways**. *Genome biology*. 2010;11(1):R3. doi:[10.1186/gb-2010-11-1-r3](https://doi.org/10.1186/gb-2010-11-1-r3).
- Nikitin A, Egorov S, Daraselia N, Mazo I. **Pathway studio—the analysis and navigation of molecular networks**. *Bioinformatics*. 2003;19(16):2155-7.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. **Gene ontology: tool for the unification of biology**. *The Gene Ontology Consortium*. *Nature genetics*. 2000;25(1):25-9. doi:[10.1038/75556](https://doi.org/10.1038/75556).
- Petersen TN, Brunak S, von Heijne G, Nielsen H. **SignalP 4.0: discriminating signal peptides from transmembrane regions**. *Nature methods*. 2011;8(10):785-6. doi:[10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701).
- Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E et al. **PSICQUIC and PSIScore: accessing and scoring molecular interactions**. *Nature methods*. 2011;8(7):528-9. doi:[10.1038/nmeth.1637](https://doi.org/10.1038/nmeth.1637).
- Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. **Cytoscape Web: an interactive web-based network browser**. *Bioinformatics*. 2010;26(18):2347-8. doi:[10.1093/bioinformatics/btq430](https://doi.org/10.1093/bioinformatics/btq430).
- Smoot ME, Ono K, Ruschekinski J, Wang PL, Ideker T. **Cytoscape 2.8: new features for data integration and network visualization**. *Bioinformatics*. 2011;27(3):431-2. doi:[10.1093/bioinformatics/btq675](https://doi.org/10.1093/bioinformatics/btq675).
- Kaspric N, PB, Reichstadt M., Tournayre J., Bonnet M., editor. **Protein function easily investigated by genomics data mining using the ProteINSIDE web service**. *Proceeding of the 2nd International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO)*. 2014.
- Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. **QuickGO: a web-based tool for Gene Ontology searching**. *Bioinformatics*. 2009;25(22):3045-6. doi:[10.1093/bioinformatics/btp536](https://doi.org/10.1093/bioinformatics/btp536).
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence**. *J Mol Biol*. 2000;300(4):1005-16. doi:[10.1006/jmbi.2000.3903](https://doi.org/10.1006/jmbi.2000.3903).
- Orchard S, Ammari M, Aranda B, Breuzal L, Briganti L, Broackes-Carter F et al. **The MintAct project—IntAct as a common curation platform for 11 molecular interaction databases**. *Nucleic Acids Res*. 2014;42(Database issue):D358-63. doi:[10.1093/nar/gkt1115](https://doi.org/10.1093/nar/gkt1115).
- UniProt C. **Activities at the Universal Protein Resource (UniProt)**. *Nucleic Acids Res*. 2014;42(11):7486. doi:[10.1093/nar/gku469](https://doi.org/10.1093/nar/gku469).
- Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D et al. **The BioGRID interaction database: 2015 update**. *Nucleic Acids Res*. 2015;43(Database issue):D470-8. doi:[10.1093/nar/gku1204](https://doi.org/10.1093/nar/gku1204).
- Orchard S, Kerrien S, Abbani S, Aranda B, Bhatte J, Bidwell S et al. **Protein interaction data curation: the International Molecular Exchange (IMEx) consortium**. *Nature methods*. 2012;9(4):345-50. doi:[10.1038/nmeth.1931](https://doi.org/10.1038/nmeth.1931).
- Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stumpflen V et al. **The minimum information required for reporting a molecular interaction experiment (MIMIx)**. *Nature biotechnology*. 2007;25(8):894-8. doi:[10.1038/nbt1324](https://doi.org/10.1038/nbt1324).
- Coordinators NR. **Database resources of the National Center for Biotechnology Information**. *Nucleic Acids Res*. 2014;42(Database issue):D7-17. doi:[10.1093/nar/gkt1146](https://doi.org/10.1093/nar/gkt1146).
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S et al. **AmiGO: online access to ontology and annotation data**. *Bioinformatics*. 2009;25(2):288-9. doi:[10.1093/bioinformatics/btn615](https://doi.org/10.1093/bioinformatics/btn615).
- Al-Shahrour F, Diaz-Uriarte R, Dopazo J. **FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes**. *Bioinformatics*. 2004;20(4):578-80. doi:[10.1093/bioinformatics/btg455](https://doi.org/10.1093/bioinformatics/btg455).
- Benjamini Y, Hochberg Y. **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing**. *J Roy Stat Soc B Met*. 1995;57(1):289-300.
- Tore O. **Structure and Evolution of Weighted Networks**. 2009.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **Basic local alignment search tool**. *J Mol Biol*. 1990;215(3):403-10. doi:[10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E et al. **ExPASy: SIB bioinformatics resource portal**. *Nucleic Acids Res*. 2012;40(Web Server issue):W597-603. doi:[10.1093/nar/gks400](https://doi.org/10.1093/nar/gks400).

35. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigan C et al. **The Bioperl toolkit: Perl modules for the life sciences.** *Genome research.* 2002;12(10):1611-8. doi:[10.1101/gr.361602](https://doi.org/10.1101/gr.361602).
36. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C et al. **The IntAct molecular interaction database in 2012.** *Nucleic Acids Res.* 2012;40(Database issue):D841-6. doi:[10.1093/nar/gkr1088](https://doi.org/10.1093/nar/gkr1088).
37. Chatri-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C et al. **The BioGRID interaction database: 2013 update.** *Nucleic Acids Res.* 2013;41(Database issue):D816-23. doi:[10.1093/nar/gks1158](https://doi.org/10.1093/nar/gks1158).
38. Huntley RP, Binns D, Dimmer E, Barrell D, O'Donovan C, Apweiler R. **QuickGO: a user tutorial for the web-based Gene Ontology browser.** *Database : the journal of biological databases and curation.* 2009;2009:bap010. doi:[10.1093/database/bap010](https://doi.org/10.1093/database/bap010).
39. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nat Protoc.* 2007;2(4):953-71. doi:[10.1038/nprot.2007.131](https://doi.org/10.1038/nprot.2007.131).
40. Caccia D, Dugo M, Callari M, Bongarzone I. **Bioinformatics tools for secretome analysis.** *Biochim Biophys Acta.* 2013;1834(11):2442-53. doi:[10.1016/j.bbapap.2013.01.039](https://doi.org/10.1016/j.bbapap.2013.01.039).
41. Nickel W. **The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes.** *Eur J Biochem.* 2003;270(10):2109-19.
42. Marthia VS, Liu Z, Guo L, Su Z, Ye Y, Fang H et al. **Constructing a robust protein-protein interaction network by integrating multiple public databases.** *BMC Bioinformatics.* 2011;12 Suppl 10:S7. doi:[10.1186/1471-2105-12-S10-S7](https://doi.org/10.1186/1471-2105-12-S10-S7).
43. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D. **A protein interaction network associated with asthma.** *Journal of theoretical biology.* 2008;252(4):722-31. doi:[10.1016/j.jtbi.2008.02.011](https://doi.org/10.1016/j.jtbi.2008.02.011).
44. Blake JA, Bult CJ. **Beyond the data deluge: data integration and bio-ontologies.** *Journal of biomedical informatics.* 2006;39(3):314-20. doi:[10.1016/j.jbi.2006.01.003](https://doi.org/10.1016/j.jbi.2006.01.003).
45. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W et al. **Big data: The future of biocuration.** *Nature.* 2008;455(7209):47-50. doi:[10.1038/455047a](https://doi.org/10.1038/455047a).
46. Gobeill J, Pasche E, Vishnyakova D, Ruch P. **Managing the data deluge: data-driven GO category assignment improves while complexity of functional annotation increases.** *Database-Oxford.* 2013. doi:[10.1093/database/bat041](https://doi.org/10.1093/database/bat041).
47. Carnielli CM, Winck FV, Paes Leme AF. **Functional annotation and biological interpretation of proteomics data.** *Biochim Biophys Acta.* 2015;1854(1):46-54. doi:[10.1016/j.bbapap.2014.10.019](https://doi.org/10.1016/j.bbapap.2014.10.019).
48. Huang da W, Sherman BT, Lempicki RA. **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res.* 2009;37(1):1-13. doi:[10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923).
49. Maere S, Heymans K, Kuiper M. **BINGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks.** *Bioinformatics.* 2005;21(16):3448-9. doi:[10.1093/bioinformatics/bti551](https://doi.org/10.1093/bioinformatics/bti551).
50. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. **GORilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics.* 2009;10:48. doi:[10.1186/1471-2105-10-48](https://doi.org/10.1186/1471-2105-10-48).
51. Rice P, Longden I, Bleasby A. **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends in genetics : TIG.* 2000;16(6):276-7.
52. Hiller K, Grote A, Scheer M, Munch R, Jahn D. **PrediSi: prediction of signal peptides and their cleavage positions.** *Nucleic Acids Res.* 2004;32(Web Server issue):W375-9. doi:[10.1093/nar/gkh378](https://doi.org/10.1093/nar/gkh378).
53. Kall L, Krogh A, Sonnhammer EL. **Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server.** *Nucleic Acids Res.* 2007;35(Web Server issue):W429-32. doi:[10.1093/nar/gkm256](https://doi.org/10.1093/nar/gkm256).
54. Hernandez-Toro J, Prieto C, De las Rivas J. **APID2NET: unified interactome graphic analyzer.** *Bioinformatics.* 2007;23(18):2495-7. doi:[10.1093/bioinformatics/btm373](https://doi.org/10.1093/bioinformatics/btm373).
55. Tarcea VG, Weymouth T, Ade A, Bookvich A, Gao J, Mahavisno V et al. **Michigan molecular interactions r2: from interacting proteins to pathways.** *Nucleic Acids Res.* 2009;37(Database issue):D642-6. doi:[10.1093/nar/gkn722](https://doi.org/10.1093/nar/gkn722).
56. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E et al. **MINT, the molecular interaction database: 2012 update.** *Nucleic Acids Res.* 2012;40(Database issue):D857-61. doi:[10.1093/nar/gkr930](https://doi.org/10.1093/nar/gkr930).
57. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A et al. **STRING v9.1: protein-protein interaction networks, with increased coverage and integration.** *Nucleic Acids Res.* 2013;41(Database issue):D808-15. doi:[10.1093/nar/gks1094](https://doi.org/10.1093/nar/gks1094).
58. Zhang QC, Petrey D, Garzon JI, Deng L, Honig B. **PrePPI: a structure-informed database of protein-protein interactions.** *Nucleic Acids Res.* 2013;41(Database issue):D828-33. doi:[10.1093/nar/gks1231](https://doi.org/10.1093/nar/gks1231).
59. Huang DW, Sherman BT, Lempicki RA. **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc.* 2009;4(1):44-57. doi:[10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211).
60. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A et al. **Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling.** *Nucleic Acids Res.* 2010;38(Web Server issue):W210-3. doi:[10.1093/nar/gkq388](https://doi.org/10.1093/nar/gkq388).
61. McCarthy FM, Gresham CR, Buza TJ, Chouvarine P, Pillai LR, Kumar R et al. **AgBase: supporting functional modeling in agricultural organisms.** *Nucleic Acids Res.* 2011;39(Database issue):D497-506. doi:[10.1093/nar/gkq1115](https://doi.org/10.1093/nar/gkq1115).

ENDNOTES

^a <http://www.cbs.dtu.dk/services/SignalP/performance.php>

^b <http://www.cbs.dtu.dk/services/TargetP/instructions.php>

Conclusion de l'étude

Pour la liste créée avec les 133 ID de protéines, les résultats obtenus sont globalement les mêmes pour les deux versions de ProteINSIDE, les différences résultant essentiellement des mises à jour plus récentes des bases de données. L'ajout des nouvelles fonctionnalités de ProteINSIDE a cependant augmenté la fiabilité des analyses effectuées avec l'outil. L'ajout de l'enrichissement des termes GO a révélé immédiatement les fonctions surreprésentées attendues pour le jeu de données, à savoir : « extracellular region », « glycolytic process » et « hormone activity ». Les scores de prédiction des protéines sécrétées rassurent l'utilisateur sur chaque prédiction et indiquent leurs fiabilités. Les PPI sont identifiées en plus faible quantité, car les doublons sont évités et les PPI prédites sont exclues.

Les mises à jour de l'interface de ProteINSIDE rendent son utilisation plus conviviale. Par exemple, les réseaux sont plus simples à trier, ce qui en améliore leurs analyses. Les résultats s'affichent plus vite et sont immédiatement réutilisables pour des chercheurs non-programmeurs. La rapidité générale de l'outil a été grandement augmentée grâce à sa migration sur un serveur plus puissant.

1.4 – Conclusion générale à propos de l'outil ProteINSIDE

ProteINSIDE est un outil en ligne constitué de plusieurs modules dédiés à la fouille automatique des données de génomique. Il est destiné à analyser les données des ruminants (bovin, mouton et chèvre), mais aussi issues d'espèces monogastriques (homme, souris et rat). ProteINSIDE simplifie l'interprétation des résultats grâce à une interface web interactive et à des résultats téléchargeables à des formats communément utilisés.

Une comparaison des résultats produits par les deux versions de ProteINSIDE montre des différences ([Tableau 8](#)). Une synthèse des résultats obtenus avec une analyse basique (paramètres par défaut) et une analyse personnalisée (identique à l'analyse basique, mais en ajoutant des annotations électroniques et des interactions entre les protéines du jeu de données et des protéines extérieures), réalisées avec la version bêta de ProteINSIDE et avec la version actuellement disponible, sont présentés dans le [Tableau 8](#). Les résultats concernent le jeu de données construit avec les 133 ID, décrit précédemment (cf. paragraphe 1.2). La version actuelle de ProteINSIDE annote 128 protéines par plus de termes GO, ceci peut s'expliquer par les mises à jour des bases de données de GO entre les deux versions. Cependant, l'ajout de l'enrichissement des termes GO montre que seuls certains termes sont significatifs pour ce jeu de données. Les résultats sont similaires entre les deux versions pour la prédiction de protéines sécrétées avec un peptide signal, en revanche les confirmations de sécrétions par les termes GO sont plus restreintes. Le dictionnaire des termes GO liés à une fonction de sécrétion est plus précis donc moins de GO confirment les résultats de prédiction.

Les résultats de localisations cellulaires confirment que 82 des 85 protéines prédites comme sécrétées le sont. Parmi elles, 16 protéines sont prédites comme potentiellement sécrétées avec de forts scores de prédiction (selon SignalP et TargetP ; 22 pour l'analyse personnalisée). Dans la version actuelle de ProteINSIDE, la recherche de protéines potentiellement sécrétées sans peptide

signal ajoute une trentaine de protéines comme potentiellement sécrétées par d'autres voies de sécrétion. Parmi ces 30 protéines, on dénombre 16 protéines prédites par TargetP comme potentiellement sécrétées par une voie de sécrétion autre que celle impliquant un peptide signal.

Les PPI identifiées dans l'analyse basique sont moins nombreuses dans la version bêta que dans la version actuelle de ProteINSIDE, mais elles sont effectuées par plus de protéines du jeu de données. Ceci résulte d'une gestion différente de la sélection des PPI selon les méthodes de détection. En effet, des méthodes de détection étaient écrites différemment selon les bases de données source : la méthode « two hybrid », par exemple, était écrite « 2h », « 2 H », « 2 hybrid » ou encore « two-hybrid-test ». Pourtant, il s'agit bien de la même méthode et les doublons étaient pris en compte si deux protéines interagissaient avec cette méthode écrite de deux façons différentes. Le nombre de PPI identifiées dans l'analyse paramétrée est plus important dans la version actuelle que dans la version bêta de ProteINSIDE et implique plus de protéines, ceci s'explique par les mises à jour constantes des bases de données d'interactions. Ainsi, les résultats obtenus pour ce jeu de données test sont conformes aux résultats attendus en terme d'annotations GO surreprésentées, de nombre de protéines prédites comme sécrétées et de PPI.

Modules	Version bêta de ProteINSIDE		Version actuelle de ProteINSIDE	
	Analyse basique	Analyse complète	Analyse basique	Analyse complète
ID Mapping	132	132	132	132
Gene Ontology	568 termes GO annotent 123 protéines	1031 termes GO annotent les 132 protéines	- 624 termes GO annotent 128 protéines - 114 termes GO avec des p -value < 0,0001	1080 termes GO annotent les 132 protéines - 238 termes GO avec des p -value < 0,0001
Secreted Proteins	85 prédictions de SignalP dont 63 confirmées par des termes GO	85 prédictions de SignalP dont 82 confirmées par des termes GO	- 85 prédictions de SignalP dont 82 confirmées par les localisations cellulaires et 25 par des termes GO - 30 protéines par des voies de sécrétion n'impliquant pas de peptide signal	85 prédictions de SignalP dont 82 confirmées par des termes GO - 32 protéines par d'autres voies de sécrétion
Protein Interaction	36 interactions par 23 protéines du jeu de données	612 interactions par 221 protéines, dont 61 du jeu de données	29 interactions par 28 protéines du jeu de données	688 interactions par 500 protéines, dont 61 du jeu de données

Tableau 8 : Résultats produits par la version bêta et la version aboutie de ProteINSIDE.

2 – Mise en application de ProteINSIDE

2.1 – Introduction

Nous avons précédemment décrit les performances techniques de ProteINSIDE sur un jeu de données de 133 ID que nous avons composé. Nous avons également montré qu'il existe des outils bioinformatiques performants pour la fouille de données de génomique pour les espèces modèles ou l'homme, mais pas forcément adaptés aux espèces agronomiques. Ainsi, pour **démontrer la fiabilité et les apports originaux de ProteINSIDE, il paraissait indispensable de comparer les résultats produits par ProteINSIDE à ceux produits par d'autres outils utilisés par la communauté scientifique.** Pour cela, nous avons confronté ProteINSIDE à DAVID (l'outil d'analyse de données de génomique le plus utilisé et référencé actuellement dans la littérature pour les annotations GO essentiellement ; ([Huang et al., 2009](#))), à BioMyn (un outil d'analyse de listes de protéines ou de gènes humains dont les modules d'analyses sont similaires à ceux de ProteINSIDE ; ([Ramirez et al., 2012](#))) et à AgBase (une base de données qui possède aussi des outils pour la synthèse de données et l'annotation fonctionnelle par des termes GO, tout en étant dédiée aux espèces agronomiques dont les bovins et les ovins ([McCarthy et al., 2011](#))). Pour tester les performances de ProteINSIDE comparativement à ces 3 outils, nous avons créé six ensembles de données composés de 1000 protéines choisies aléatoirement pour chacune des six espèces. Ces protéines ont été sélectionnées en utilisant un script qui prend au hasard 1000 protéines sur la base de données UniProt et ceci pour chaque espèce concernée dans ProteINSIDE.

Dans un deuxième temps, la pertinence biologique des résultats produits par ProteINSIDE a été testée par l'analyse de données de protéomique. Ces données proviennent de précédentes études publiées pour le muscle *Semitendinosus* ([Chaze et al., 2008](#); [Chaze et al., 2009](#)) et le TA périrénal ([Taga et al., 2012](#)) de fœtus bovins de race Charolaise et Blonde d'Aquitaine. Ces résultats correspondent à deux listes de 143 et 120 protéines, respectivement pour le TA et pour le muscle, identifiées comme présentes dans les tissus à 60 (uniquement pour le muscle), 110, 180, 210, et 260 jours post-conception (jpc) ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)). Les premières fouilles de données pour ces listes de protéines ont précisé l'implication des protéines identifiées dans des processus de prolifération et de différenciation accompagnant la croissance des deux tissus. Cependant, les données des deux tissus ont été analysées indépendamment et la constitution d'un jeu de données commun aux deux listes pourrait mettre en évidence des processus biologiques partagés lors de la croissance fœtale. Aussi, **nous avons constitué un jeu de données des protéines communes aux deux tissus afin d'apporter de nouvelles connaissances sur l'ontogenèse des TA et des muscles et sur les interactions entre ces deux processus.** Cette étude est rendue possible grâce à l'augmentation des annotations disponibles depuis trois ans et à une fouille de données plus complète à l'aide de ProteINSIDE.

ARTICLE 3

ProteINSIDE to Easily Investigate Proteomics Data from Ruminants:
Application to Mine Proteome of Adipose and Muscle Tissues in
Bovine Foetuses

Nicolas KASPRIC^{1, 2}, Brigitte PICARD^{1, 2}, Matthieu REICHSTADT^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213, Lempdes, France

PLOS ONE, doi:10.1371/journal.pone.0128086.– Publié le 22 mai 2015 (24 pages).

RESEARCH ARTICLE

ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses

Nicolas Kaspric^{1,2}, Brigitte Picard^{1,2}, Matthieu Reichstadt^{1,2}, Jérémy Tournayre^{1,2}, Muriel Bonnet^{1,2*}

1 UMR1213 Herbivores, INRA, Saint-Genès-Champanelle, France, **2** UMR1213 Herbivores, Clermont Université, VetAgro Sup, Clermont-Ferrand, France

* muriel.bonnet@clermont.inra.fr


 OPEN ACCESS

Citation: Kaspric N, Picard B, Reichstadt M, Tournayre J, 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

Academic Editor: Frederique Lisacek, Swiss Institute of Bioinformatics, SWITZERLAND

Received: February 6, 2015

Accepted: April 23, 2015

Published: May 22, 2015

Copyright: © 2015 Kaspric et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the regional council of Auvergne in France and APIS-GENE. The funders had no role in study design, data collection and analysis, or preparation of the manuscript. APIS-GENE gave the authors its agreement to publish.

Competing Interests: All authors are in agreement with the content of the manuscript. The authors declare that they have no financial and personal

Abstract

Genomics experiments are widely acknowledged to produce a huge amount of data to be analysed. The challenge is to extract meaningful biological context for proteins or genes which is currently difficult because of the lack of an integrative workflow that hinders the efficiency and the robustness of data mining performed by biologists working on ruminants. Thus, we designed ProteINSIDE, a free web service (www.proteinside.org) that (I) provides an overview of the biological information stored in public databases or provided by annotations according to the Gene Ontology, (II) predicts proteins that are secreted to search for proteins that mediate signalisation between cells or tissues, and (III) analyses protein-protein interactions to identify proteins contributing to a process or to visualize functional pathways. Using lists of proteins or genes as a unique input, ProteINSIDE is an original all-in-one tool that merges data from these searches to present a fast overview and integrative analysis of genomic and proteomic data from Bovine, Ovine, Caprine, Human, Rat, and Murine species. ProteINSIDE was bench tested with 1000 proteins identifiers from each species by comparison with DAVID, BioMyn, AgBase, PrediSi, and Phobius. Compared to DAVID or BioMyn, identifications and annotations provided by ProteINSIDE were similar from monogastric proteins but more numerous and relevant for ruminants proteins. ProteINSIDE, thanks to SignalP, listed less proteins potentially secreted with a signal peptide than PrediSi and Phobius, in agreement with the low false positive rate of SignalP. In addition ProteINSIDE is the only resource that predicts proteins secreted by cellular processes that do not involve a signal peptide. Lastly, we reported the usefulness of ProteINSIDE to bring new biological hypotheses of research from proteomics data: the biological meaning of the uptake of adiponectin by the foetal muscle and a role for autophagy during ontogenesis of adipose and muscle tissues.

relationships with other people or organizations that can inappropriately influence their work.

Introduction

A main challenge for scientists working on the efficiency of ruminant production and the quality of their products (meat, milk. . .) is to understand which genes and proteins control nutrient metabolism and partitioning between tissues or which genes and proteins control tissues growth and physiology [1–3]. These researches have produced vast amount of proteomics and genomics data and have used bioinformatics analyses to extract meaningful biological context for proteins or genes in ruminants [4–10]. Therefore, available bioinformatics tools implemented as web services are used to address these issues such as Amigo [11], Gorilla [12], or QuickGO [13] dedicated to Gene Ontology (GO) annotation, GeneCards [14] for an overview of the current available information about a gene or a protein for Human species, MiMi [15], IntAct [16], BioGrid [17], or PsicQuick [18] for protein-protein interactions (PPI) identification, STRING [19] or Apid2Net [20] for PPI identification and visualization as networks, and Sigcleave [21], SignalBlast [22], or SignalP [23] for the prediction of signal peptides. To gain insight efficiency and robustness, free and commercial tools such as DAVID [24], BioMyn [25], ToppGene [26], Expander [27], PathwayStudio [28], or the web-based software package Ingenuity Pathway Analysis (<http://www.ingenuity.com>; Redwood City, CA, USA) were developed as one-stop portal in order to combine two or three analyses, mainly GO annotations, PPI interaction, and information retrieval. Most of them do not permit to address data mining from genomic and proteomic studies in ruminants, except AgBase [29], a curated genomic database providing functional annotations and information for agriculturally important species, among them bovine and ovine. None of these resources include the prediction of proteins that are secreted outside the cell as recommended for a computational analysis of secretome [30]. However, in the context of livestock production, the efficiency of animals or the quality of their products not only depend on the functioning of a tissue but also on the interactions between tissues or cell types within a tissue [1], partly mediated by secreted proteins. Secreted proteins play important roles in tissue growth and homeostasis, as well as in tissues cross-talk and nutrient partitioning [2,31].

Here, we present ProteINSIDE, which aims to seamlessly integrate complementary analyses to produce information about proteins in one online automated package. ProteINSIDE comprises four modules: biological knowledge retrieval, annotations relative to biological process, molecular function, and subcellular location according to the GO, prediction of secreted proteins, and PPI analysed as a network. ProteINSIDE provides graphical and interactive results viewable on the website or downloadable. ProteINSIDE extracts information for a list of genes or proteins ID from myriad data sources with a unique input. Thus, it will circumvent searching for the different ID required for queries with the multiple existing bioinformatics tools, and it will save analysis time for the biologists. We demonstrated the higher or similar performances of ProteINSIDE relatively to currently used web-services or dedicated resources by a bench test that has evaluated results from 1000 random proteins by species. The relevance of results produced by ProteINSIDE was also checked with data that were previously partly analysed [3–5]. Lastly, we report the new knowledge related to adipogenesis, myogenesis, and the balance between both processes, provided by mining dataset from bovine foetal muscle and adipose tissues (AT) thanks to ProteINSIDE.

Materials and Methods

1 ProteINSIDE features

ProteINSIDE is an online workflow with an interface devoted to user-friendly and fully customizable analyses from lists of proteins or genes ID. Registered users have access to a private

session to run, save, and visualise their results. Unregistered users can use ProteINSIDE, there is no analyses manager and results are deleted each month. Uploaded data are encrypted to ensure confidentiality. ProteINSIDE is divided into three parts: the workflow, the database, and the web interface. The workflow is a combination of Perl and R scripts to query databases, recover protein data, perform calculations, and run algorithms for signal peptide predictions and network visualisation. The MySQL database aims to reduce server load and stores settings and results from queries. ProteINSIDE's database also stores available knowledge from major public biological databases. The web interface is the structure of ProteINSIDE and allows creating an analysis, viewing results, and keeping users informed with updates (Fig 1).

2 Setting up a basic or a custom analysis/query

ProteINSIDE performs either a "basic analysis" (in which settings are locked and the workflow acquires GO terms, signal peptide prediction, and PPI data from IntAct, UniProt, and BioGrid

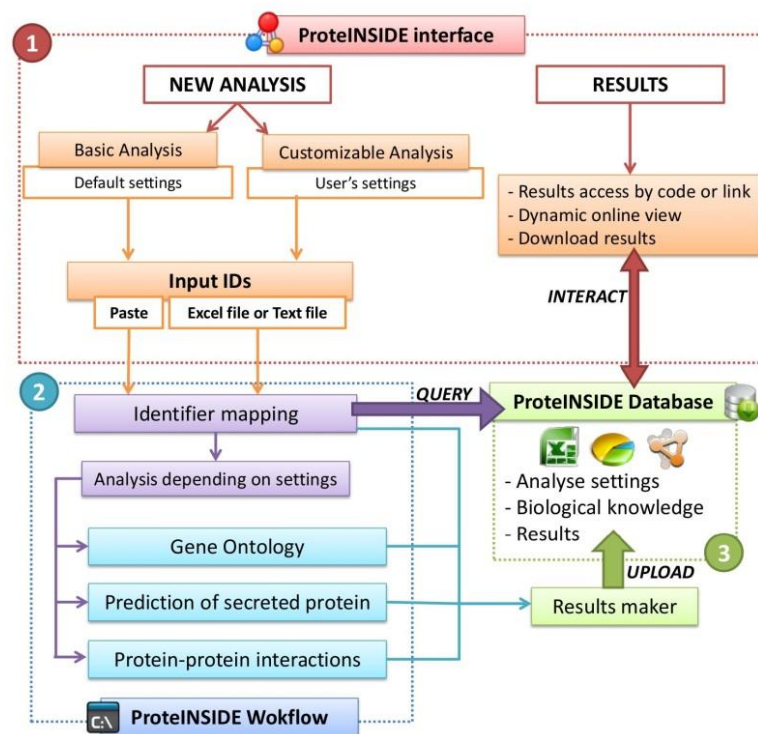


Fig 1. Flowchart of ProteINSIDE structure. The four modules to query the available biological information, annotate according to the GO, predict secreted proteins and visualize PPI, are either all run in the basic analysis or individually selected and run with specific settings in the custom analysis. The basic analysis runs ProteINSIDE with automatic settings. The custom analysis operates with the settings selected by the user: option to include GO Inferred from IEA codes (electronic annotation that are automatically unselected in the basic analysis), option to make GOTree chart networks with Cytoscape web, option to search PPI among the 31 databases proposed by ProteINSIDE, option to search PPI in other species using orthologous proteins, option to extend the PPI network with proteins that are not in the dataset, and option to choose the sensitivity to detect signal peptides with SignalP 4.1.

doi:10.1371/journal.pone.0128086.g001

respectively) or a “custom analysis” (in which the user chooses the modules, the settings, and databases for PPI) (Fig 1). Previously defined settings for a custom analysis can be used for new queries by registered users using the “preset analysis”. To submit a basic or a custom analysis, users either directly paste a list of proteins or genes ID, or upload a file. Inputs can be protein (e.g., ADIPO_HUMAN) or gene name (e.g., ADIPO), protein accession number (e.g., Q15848), or also gene identifier (e.g., gj|62022275) from six species: Bovine, Human, Rat, Mouse, Ovine, and Caprine. A new analysis is run directly or is placed on a waiting list if the workflow is overloaded.

2.1 Biological knowledge retrieval. Submitted ID, names, or accession numbers are compared to the ProteINSIDE database to ascertain a match with genes or proteins from Human, Rat, Murine, Bovine, Ovine, or Caprine species. This is achieved by merging the local database, which is updated every month, with data from UniProt [32] and NCBI (Gene, Proteins, and HomoloGene; [33]) public databases for all known genes and proteins from these six species.

2.2 Functional annotation according to GO. ProteINSIDE imports GO terms by querying the QuickGO database. QuickGO was chosen because of its daily update, accessibility and performances. In the basic analysis, ProteINSIDE only imports GO terms that have been selected by evidence codes (GO terms Inferred from Electronic Annotation (IEA) are excluded) and agreed by curator review. The use of IEA is a setting of the custom analysis. ProteINSIDE provides the number of genes or proteins annotated by a specific GO term relatively to the total number of genes or proteins within both the dataset (frequency within the list) and the GO (frequency within the genome). The GO module of ProteINSIDE also analyses over- and under-represented terms to identify the most relevant and the most specific terms associated with the uploaded list, according to the functional enrichment first proposed by FatiGO [34]. Briefly, the frequency of a specific GO within the dataset is compared to its frequency in the genome. For this, ProteINSIDE prepares a 2 by 2 contingency table for a specific GO, performs a Fisher exact test and reports significant over-expressed GO based on the p-value of the test. The test is performed by an R script as follow:

```
mat <- matrix(c(A, B, C, D), nrow = 2)

tes <- fisher.test(mat)
```

where A is the number of proteins annotated by a specific GO term in the dataset, B the total number of proteins annotated in the dataset, C the number of proteins that are annotated by the specific GO term within the genome (also called “background of a GO term”), and D the total number of annotated proteins in the genome. ProteINSIDE provides the significantly over-expressed GO based on both unadjusted p-value of the test and a p-value corrected for multiple testing correction by the FDR Benjamini & Hochberg (BH) method [35].

The GO module also provides a view of network that links GO terms as a tree ancestor. The GO terms are linked by their parental association using ProteINSIDE database. ProteINSIDE database has its own version of ontology annotation from GO consortium [36] and QuickGO databases that relates: GO terms, biological functions, GO terms associations (relationships between terms only include “is_a” or “parts of” to select terms with a direct link), and the number of annotations by species. The “GOTree” view of ProteINSIDE is dynamic and only includes linked GO terms for a list of ID. The network is built with a custom version of Cytoscape web (v. 1.0.4) [37].

2.3 Prediction of secreted proteins. In eukaryotes, at least five different routes of protein secretion out of the cells are reported: (I) the classical Golgi/ER-dependent secretory pathway for proteins that contain N-terminal signal peptide and non-classical protein exports or ER/

Golgi-independent protein secretions that ensure protein secretion by (II) endosomal recycling, (III) plasma membrane transporter, (IV) membrane flip-flop, and (V) membrane blebbing that involves formation of vesicles or exosomes [38].

To identify proteins that are putatively secreted by the classical Golgi/ER-dependent secretory pathway, ProteINSIDE predicts the presence of a signal peptide on a protein sequence (imported by the biological knowledge retrieval script) through a local version of the SignalP tool (version 4.1 [23]). SignalP was chosen because of its high prediction score in comparison with other available tools [30,39] and the low false positive rate of approximately 6% (supplementary materials and methods of [23]). ProteINSIDE uses in the basic analysis the default cutoff value of 0.45 recommended by SignalP for a reliable prediction (optimized for correlation). A “sensitive” option is available in the custom analysis to increase the number of detected signal peptides (the cutoff value is decreased to 0.34), however with more false positive results (according to SignalP tutorial: <http://www.cbs.dtu.dk/services/SignalP/performance.php>). Then, ProteINSIDE checks that predicted secreted proteins are also annotated with GO terms related to secretion. For this purpose, we selected about 1200 GO terms related to secretion terms (monthly updated) as for example: secretion, vesicle, or also extracellular region. To identify proteins that are putatively secreted by non-classical protein exports or ER/Golgi-independent protein secretions, ProteINSIDE checks the subcellular location of proteins and compares their annotation relatively to our list of GO terms related to secretion. Lastly, to support these predictions (classical and non-classical ways) ProteINSIDE runs the TargetP software [40] that predicts the subcellular location of the proteins using the amino acid sequence. TargetP uses 4 cutoffs for its predictions: chloroplast (only for plant input), mitochondrion, secretory pathway (signal peptide), and other subcellular location. ProteINSIDE uses a preset cutoff option to get a significant prediction higher than 95% (according to TargetP tutorial: <http://www.cbs.dtu.dk/services/TargetP/instructions.php>).

2.4 Protein-protein interactions and visualisation. PPI identification and visualisation within a network point out how various genes or proteins contribute to cellular or metabolic processes. ProteINSIDE uses PsicQuic web service [18] to identify PPI and imports identified PPI by their “interaction detection methods” experimentally proven and agreed by curator review. Basic or custom analyses identify PPI within the uploaded dataset using the major preselected databases (IntAct, UniProt, and BioGrid chosen for the high data curation) or databases selected by the user (among 31 suggested by the PsicQuic web service). ProteINSIDE favours the use of multiple PPI databases in order to provide a more complete PPI network. In the custom analysis, ProteINSIDE identifies interactions between proteins from the dataset (core network) and proteins outside the dataset (extended network). ProteINSIDE identifies PPI recorded either within the species of the analysis or within a different species (to be selected among the six). To choose another species, user have either to select the Blastp [41] setting to know the similarity between the sequences of uploaded proteins and their orthologs, or to use conserved proteins declared in the HomoloGene database (only for Human, Bovine, Mouse, and Rat). For the extended network, ProteINSIDE does not include links between proteins outside of the dataset themselves in order to not overload the network visualization with too many interactions. ProteINSIDE focuses only on PPI between proteins of the dataset and other proteins of the analysis species. If users wish to view and analyze a complete network with links between proteins outside the uploaded dataset, they have to download the identifiers of the PPI extension (available on the PPI table results page) and run a new analysis.

Thus, ProteINSIDE constructs networks with PPI recorded in well- (Rat, Mouse, or Human) and poorly-annotated (Bovine, Ovine, or Caprine) species. The comparison of PPI networks allows expanding knowledge in the poorly- (Bovine, Ovine, and Caprine) relatively to well-annotated (Rat, Mouse, or Human) species. After a custom analysis with PPI extension,

a GO analysis with the ID of genes/proteins of the extended network is directly launched by clicking on the button "Run a job to analyse the Gene Ontology for ID from this network". Results of this GO analysis are available as a new work on the user home page, and provide GO terms with p -values for enrichment that take into account news values (related to the higher number of proteins) for A, B, C, D as defined in the GO annotation section 2.2. The PPi modules provide a view of proteins as networks which are built with Cytoscape web: edge colours inform on the experimental method used to identify the interaction. Each edge/experimental method has its own colour in the network and an interaction between two ID could have been detected by several experimental methods (this resulting in several edges between two ID). Key proteins can be highlighted using betweenness (that quantifies how frequently a node is on the shortest path between every pair of nodes for detecting bottlenecks in a network) or closeness (that quantifies how short are minimal paths from a given node to all others, a large closeness indicates that a node is close to the topological centre of the network) centralities as previously defined by Hwang et al. [42]. These centralities were used because they were proven as efficient to reveal key proteins that play important roles in a network [42].

2.5 Results visualisation and download. Results are available through a unique code or a link provided after the submission of a list. Four separated pages provide results from the four analyses. The results are dynamic tables and charts that can be sorted and filtered online on the website by specific criterion such as biological function or protein and gene ID. Tables and charts are downloadable, and diagrams or histograms are printable. Networks are downloadable as picture (.pdf and .png) or as network software input (.sif or .xgmml or .graphml) files. Whatever the networks (GOTree or PPi), they are dynamics thanks to Cytoscape web which gives options to sort nodes, change layout, and search by proteins or biological function (Figs 2 and 3).

2.6 Implementation. The web interface was programmed in PHP, HTML, and JavaScript. The workflow has been programmed using Perl (with CPAN modules (Comprehensive Perl Archive Network) and BioPerl [43]) and R scripts. The database was made in MySQL. By using these most common web dynamic programming languages, ProteINSIDE is completely available on any operating system using an internet browser (Internet Explorer, Firefox, Chrome, Safari. . .).

To ensure the sustainability of ProteINSIDE, we have set up an automatic update of the database. A program updates each month the biological information of the database of ProteINSIDE by comparison with the current and free NCBI and UniProt databases resources. Moreover, it gathers the information from the databases QuickGO and Amigo in order to update the GO term database of ProteINSIDE (function and annotated genes expected by GO term for each species) or to add new GO terms. The workflow update is done manually. If a program requires an update, it is set up and tested locally to avoid conflicts with running analyses of users. When the update is applied to ProteINSIDE: users can press the button "reload analysis" (green arrow) to restart previous analysis and to benefit to the new version of the workflow. Each new version of ProteINSIDE's database or workflow is informed on the website by a new or by a message in the "About" section and the "User's main page". Users should be aware that results obtained with 2 different versions of ProteINSIDE may be not reproducible because of the deletion of obsolete data and the use of new ones.

3 Datasets

To test ProteINSIDE's performances, we have created six datasets made of 1000 random proteins for each species (S1 Table lists the six samples files). These proteins have been selected using a script which queries randomly 1000 reviewed proteins for a species on the UniProt

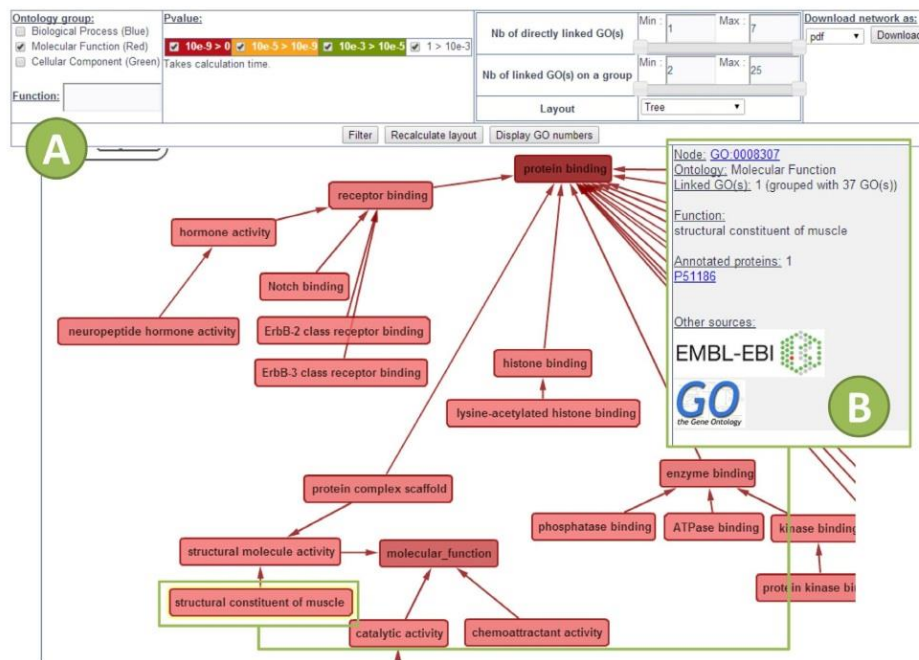


Fig 2. Dynamic GO tree network provided by ProteINSIDE. Network was built with Cytoscape web. Settings are available to sort the network according to Ontology groups, biological function, result of GO enrichment (p-value), numbers (Nb) of GO, or network layout. Go terms are sorted and colored depending on the ontology group and the number of annotated proteins. (A) Dynamic network view of GO terms related to Molecular Function. (B) Clicking on a GO term provides the GO number, proteins from the sample list annotated by this GO, and links with public GO databases (AmiGO and QuickGO).

doi:10.1371/journal.pone.0128086.g002

database. We have made separate custom analysis to test each part of the workflow using these datasets of 1000 random proteins by species. We have tested GO annotation using two custom analyses: with and without electronic annotation (option IEA). The prediction of secreted proteins has been bench tested using a custom analysis with SignalP prediction and GO annotation. Lastly, PPI were tested using four custom analyses: two analyses have queried the 3 databases (BioGrid, Uniprot, and IntAct) selected by default in the basic analysis to identify core and extended networks, and two analyses have queried 9 other databases (BioGrid, IntAct, MINT, MatrixDB, STRING, Reactome, InnateDB-IMEx, UniProt, and I2D-IMEx, among those to be chosen by the user) to identify core and extended networks.

Moreover, to demonstrate the added-value of the biological meanings provided by ProteINSIDE, we used 2 datasets that are lists of bovine proteins identified in the skeletal muscle *Semitenidosus* and the perirenal AT of Charolais and Blond d'Aquitaine fetuses [3–5]. The proteome was analysed at 60 (only for muscle), 110, 180, 210, and 260 days post conception (dpc) chosen according to previous data on muscle myogenesis [2]. As described in previous papers, all experimental procedures were approved by the Institutional Animal Care and Use Committee of INRA in accordance with the *Use of Vertebrates for Scientific Purposes Act 1985*. More recently the production of bovine fetuses was approved by the “Comité d'éthique en matière d'expérimentation animale Auvergne” (IACUC CEMEA Auvergne, agreement number CE 34–11).

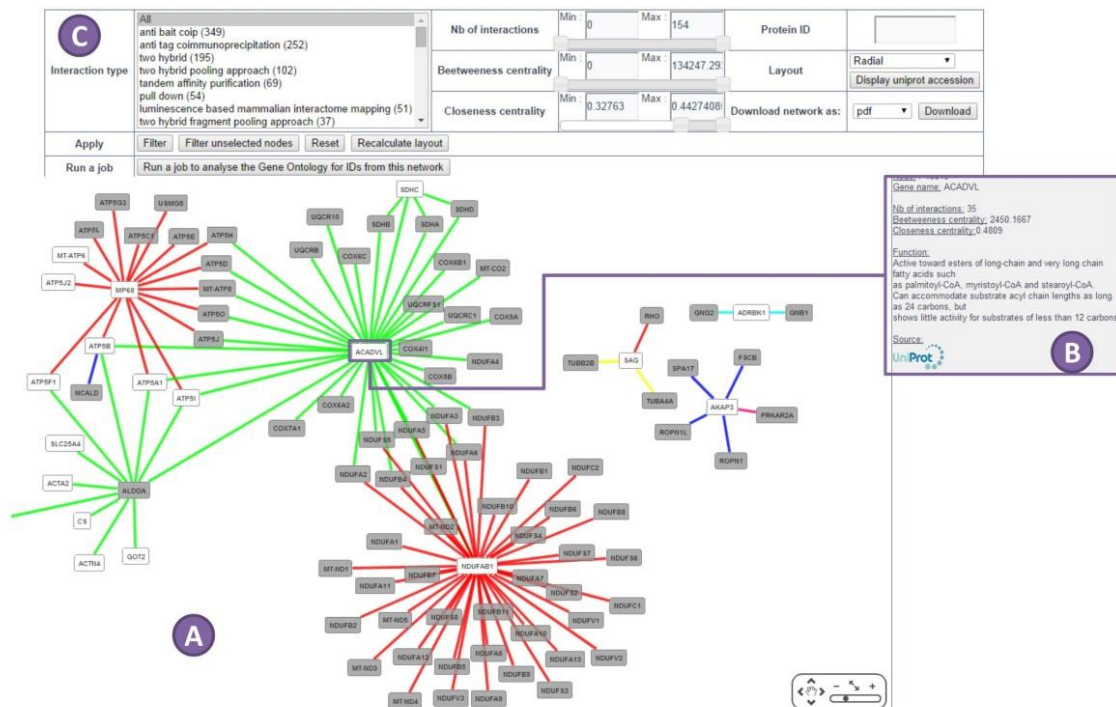


Fig 3. View of a PPI dynamic extended network made by ProteINSIDE. Network was built with Cytoscape web. (A) The colour of the edge depends on experimental methods used to identify PPI. White nodes are proteins from the dataset and grey nodes are known interacting proteins not included in the dataset. (B) Clicking on a protein/node provides biological information as gene and proteins ID, function, and a link to UniProt database. (C) Available options to sort the network and highlight proteins of interest.

doi:10.1371/journal.pone.0128086.g003

Results and Discussion

1 Evaluation of ProteINSIDE web service accuracy

We have tested ProteINSIDE with 1000 random proteins ID from six species by comparison with widely used resources (Table 1). DAVID [24] was chosen because it is a free and widely used resource (cited in more than 15500 publications, <http://scholar.google.com/citations?user=dMn7gzYAAAAA>) in several species including the three ruminant species of interest, Bovine, Ovine, and Caprine. BioMyn [25] is a free comprehensive data warehouse that unifies, integrates and explores information related to Human genes and proteins. AgBase [29] is a curated warehouse of biological information and annotations specifically for species related to agriculture, among them Ovine and Bovine. We have compared results provided with three modules of ProteINSIDE to results from DAVID, BioMyn, and AgBase web services (Tables 2–4). To our knowledge, none of these three tools or current bioinformatics integrative resources performs the prediction of secreted proteins (Table 1). Results from the secretion module of ProteINSIDE were thus compared to two specific resources dedicated to the computational prediction of protein secretion PrediSi [44] and Phobius [45], chosen for their good scoring (supplementary data of [23] and [30]).

Table 1. Available modules of analyses for ProteINSIDE, DAVID, BioMyn, and AgBase.

Analyses	ProteINSIDE	DAVID	BioMyn	AgBase
ID Mapping	x	x	x	x
<i>Protein & Gene ID</i>	x	x	x	x
<i>Biological function</i>	x	x		x
<i>Chromosomal location</i>	x	x		
<i>Tissue specificity</i>	x			x
<i>Cellular location</i>	x			
Gene Ontology	x	x	x	x
<i>Functional annotation graph</i>	x	x	x	
<i>Enrichment tests</i>	x	x	x	
<i>Expected gene products</i>	x		x	
<i>GO dynamic network</i>	x			
<i>Select/unselect IEA</i>	x			x
<i>Database source</i>	x		x	x
Prediction of secreted proteins	x			
<i>Confirmed with GO of secretion</i>	x			
<i>Confirmed with TargetP prediction</i>	x			
PPI	x	x	x	
<i>Number of queried PPI databases</i>	31	4	10	
<i>Dynamic PPI network visualization</i>	x		x	
Results				
<i>Downloadable (Excel and/or text files)</i>	x	x	x	x
<i>Online interface to view and sort results</i>	x	x	x	
ID conversion	x	x		
Orthologs conversion	x			
Species (vs. the 6 analysed by ProteINSIDE)	6	6	Human	Bovine & Sheep
Last Update (frequency)	Monthly	Sep. 2009	Mar. 2013	Monthly

Analyses performed by ProteINSIDE in comparison with DAVID [24], BioMyn [25], and AgBase [29]. ProteINSIDE performs analyses using gene or protein ID from 6 species: Human, Rat, Murine, Bovine, Ovine, and Caprine species. x indicates that the analysis is performed by the resource.

doi:10.1371/journal.pone.0128086.t001

Table 2. Comparison of identifier mapping results.

Species	ProteINSIDE	DAVID	BioMyn	AgBase
Human	1000	899	998	
Mouse	998	993		
Rat	1000	949		
Sheep	979	378		1000
Goat	1000	6		
Bovine	1000	959		1000
% of retrieval for 6 species	99.62	69.73		
% of retrieval for ruminants (%)	99.30 ^a	44.77 ^a		100 ^b

Numbers of ID retrieved by the ID mapping module of ProteINSIDE and by DAVID, BioMyn, or AgBase when a list of 1000 random proteins per species was uploaded.

^a from Sheep, Goat and Bovine results

^b from Sheep and Bovine results

doi:10.1371/journal.pone.0128086.t002

Table 3. Numbers of annotated proteins and GO terms provided by ProteINSIDE, DAVID, BioMyn, and AgBase.

Species		ProteINSIDE	DAVID	BioMyn	AgBase	Percent of results shared by ProteINSIDE and		
						DAVID	BioMyn	AgBase
Human	Annotated proteins	816 (945)	- (864)	- (803)		- (90)	- (85)	
	GO terms	2676 (3641)	- (4167)	- (1930)		- (64)	- (50)	
Mouse	Annotated proteins	938 (997)	- (991)			- (89)		
	GO terms	2435 (3514)	- (3545)			- (55)		
Rat	Annotated proteins	783 (979)	- (950)			- (91)		
	GO terms	2790 (4074)	- (5134)			- (67)		
Sheep	Annotated proteins	159 (916)	- (372)		125 (938)	- (40)		78 (100)
	GO terms	834 (2446)	- (1252)		1069 (2541)	- (51)		100 (100)
Goat	Annotated proteins	32 (503)	- (6)			- (1)		
	GO terms	190 (612)	- (82)			- (4)		
Bovine	Annotated proteins	365 (886)	- (957)		286 (898)	- (86)		77 (100)
	GO terms	1418 (3085)	- (3850)		1569 (3130)	- (61)		100 (100)

The percent of results shared by ProteINSIDE and by DAVID, BioMyn, or AgBase were calculated for comparison.

() for annotations that include IEA (Inferred electronic annotation).

- indicates that the resource does not provide annotation without IEA.

doi:10.1371/journal.pone.0128086.t003

1.1 Biological information retrieval by mapping Identifiers: ID Resume. The high ability of ProteINSIDE to retrieve biological information for 1000 random proteins ID from Human, Rat, Murine, Bovine, Ovine, or Caprine species is shown by the retrieval of 100% of ID from Bovine, Goat, Human, and Rat; as well as of 99.8% and 97.8% of ID from Mouse and Sheep, respectively (Table 2). These ID retrievals by ProteINSIDE were similar to those of BioMyn for Human ID or AgBase for Ovine and Bovine ID, but higher than those of DAVID, especially for Ovine (38% of ID retrieved) and Caprine (0.1% of ID retrieved) species (Table 2). The differences between the performances of ProteINSIDE, BioMyn or AgBASE, and DAVID

Table 4. Number of PPI identified using ProteINSIDE.

	3 databases		9 databases	
	PPI within the dataset	PPI outside the dataset	PPI within the dataset	PPI outside the dataset
Human	396	5518	1703	6269
Mouse	111	1410	702	8475
Rat	27	271	42	425
Sheep	0	0	0	0
Goat	0	0	0	0
Bovine	12	96	34	131
Bovine EXT Human			12	5921

ProteINSIDE had queried both 3 (BioGrid, Uniprot, and IntAct selected by default in the basic analysis) and 9 (BioGrid, IntAct, MINT, MatrixDB, STRING, Reactome, InnateDB-IMEx, UniProt, and I2D-IMEx chosen by the user) databases to record PPI that have been identified by experiments. Within a species, PPI were searched between proteins within (core network) and outside (extended network) the dataset. Bovine proteins ID were uploaded to search for known interactions with their orthologs in Human (EXT Human).

doi:10.1371/journal.pone.0128086.t004

may come from the lack of a recent database update for DAVID and the consequent missing of biological information (Table 1).

For each uploaded ID, ProteINSIDE obtained and summarized, as a downloadable table, the gene or protein ID, gene and protein names, the protein function, the gene chromosomal location, information on tissues expression, the cellular location, orthologous ID, and the FASTA sequence of the protein. These results are directly viewed on the “ID resume” web page of ProteINSIDE, and each protein or gene ID are linked to corresponding UniProt and NCBI web pages. A part of these biological data is also provided by DAVID, BioMyn, and AgBase.

1.2 Gene Ontology annotation. ProteINSIDE annotated proteins or genes ID with GO terms selected by evidence codes (IEA are excluded) and agreed by curator review in the basic analysis, and additionally with GO terms from IEA as a setting of the customs analysis. As ProteINSIDE, AgBase allowed unselecting IEA, while DAVID and BioMyn use IEA by default. Thus, results from the GO module of ProteINSIDE for 1000 random proteins ID were compared to DAVID and BioMyn results with IEA, and to AgBase results with and without IEA (Table 3).

On average 90% of proteins ID from Human, Rat, or Mouse annotated by ProteINSIDE were also annotated by DAVID. However, the percent of proteins annotated both by ProteINSIDE and DAVID decrease to 86%, 40%, and 0.6% for Bovine, Ovine, and Caprine ID, respectively. The numbers of GO terms from DAVID were around 52% higher for Human, Mouse, Rat, and Bovine proteins but 86% and 48% lower for Caprine and Ovine proteins, respectively, when compared to ProteINSIDE. Consequently, around 64% of GO terms from ProteINSIDE were common to those provided by DAVID for five species. On average, 85% of Human proteins were annotated both by ProteINSIDE and BioMyn with 50% of GO terms provided by the two resources. For Bovine and Ovine proteins, ProteINSIDE annotated more proteins with less GO terms than AgBase in analyses without IEA, while AgBase annotated more proteins with more GO terms when IEA annotations were used. In both analyses, GO terms provided by ProteINSIDE were all retrieved with AgBase. The annotation differences between ProteINSIDE and DAVID, or to a lesser extend BioMyn, may result from the lack of a recent update of the databases used. For example, from our lists of proteins, DAVID had provided GO terms that were declared obsolete by the GO consortium (GO:0006096 previous term was “Glycolysis”, replaced the 29th march 2014 by “Glycolytic process”) that thus could not be retrieved by resources such as ProteINSIDE and AgBase (that integrate the GO consortium updates). It is noteworthy that AgBase provided reliable GO annotations for two ruminant species, probably because of their own curated database that is also monthly updated. Thus, both AgBase and ProteINSIDE perform GO annotations for bovine and ovine, while only ProteINSIDE annotates ID from Caprine species. It is noteworthy that ProteINSIDE annotated on average 45% more proteins with 46% more GO terms when using IEA by comparison without IEA.

The results are viewed on the “GO” web page of ProteINSIDE as tables, diagrams, and GOTree charts. Unlike the other available tools, ProteINSIDE summarized main results of GO as GOTree charts which are ordered tree layout networks that link related GO terms (Fig 2).

1.3 Secreted proteins. Whatever the species, 95% of signal peptides predicted by ProteINSIDE thanks to SignalP were also predicted by PrediSi and Phobius (Fig 4) for 1000 random protein ID. However and as expected, the number of signal peptides predicted by ProteINSIDE was 49% and 24% lower than those predicted by PrediSi and Phobius, respectively. This may result from the known higher false positive rate of prediction by PrediSi and Phobius (52% and 16% according to [23], respectively) than SignalP 4.1 (6%). As an example of false positive, ALKBH5 was predicted as a secreted protein by Phobius and PrediSi but not by ProteINSIDE, while this is a nuclear protein according to UniProt database.

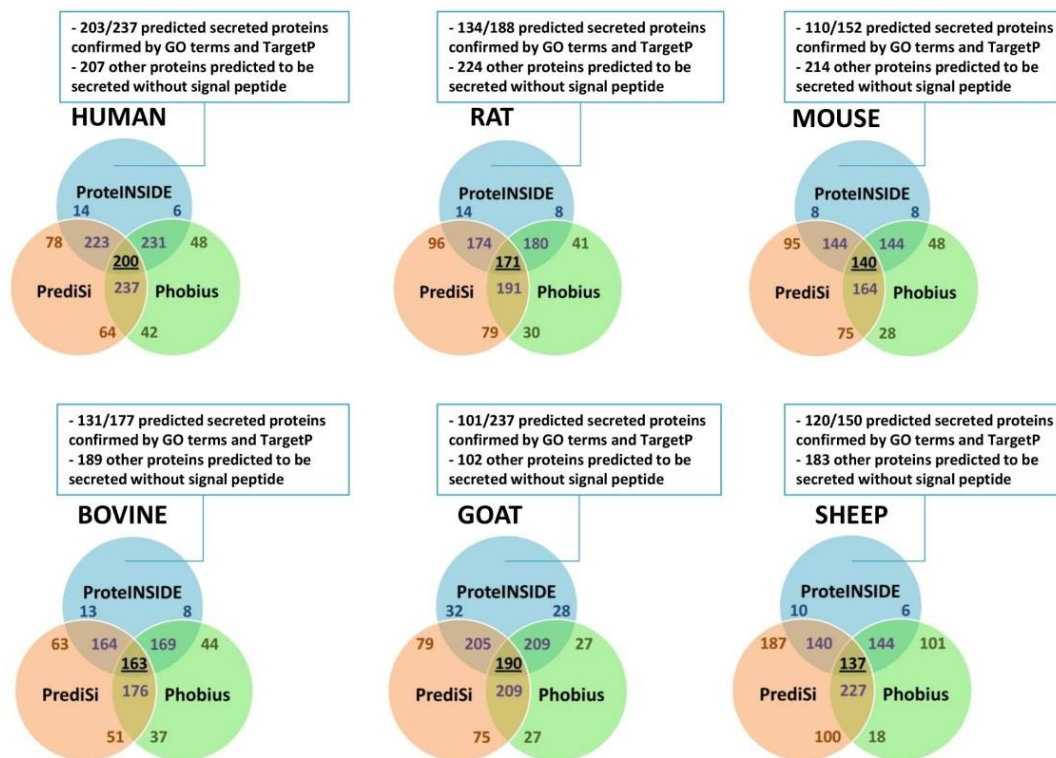


Fig 4. Venn diagrams of predicted signal peptides from amino acid sequence of 1000 proteins. Resources used to predict signal peptides were SignalP that was included in ProteINSIDE, as well as PrediSi [44] and Phobius [45] for bench tests. To reinforce the prediction of proteins to be secreted, ProteINSIDE checked the cellular localization of proteins with TargetP and used GO terms relative to the cellular component. GO terms were also used to predict proteins that are secreted without signal peptide. Squares give numbers of predicted (SignalP) secreted proteins confirmed by GO terms and subcellular location prediction (TargetP), and confirmed by GO terms for proteins that are potentially secreted without signal peptide.

doi:10.1371/journal.pone.0128086.g004

An added-value of ProteINSIDE is the use of GO terms and TargetP, both to reinforce the prediction of proteins that are secreted (Table 1). By adding these two methods, 48% (Fig 4) and 12% (in average for the six species, data not shown) of predicted secreted proteins with and without signal peptide were confirmed by GO terms and TargetP prediction, respectively. This two-step analysis improves the reliability of proteins proposed to be secreted. Moreover, unlike other available resources ProteINSIDE predicts proteins exported out of eukaryote cells by non-classical secretory pathways (without signal peptide). However, all of these predictions remain to be confirmed by new or already published experiments. Lastly, GO terms confirmation can be improved by the use of IEA codes for annotations.

Results are viewed on the “Secreted Protein” web page of ProteINSIDE as two main tables: the first table listed proteins secreted by the classical Golgi/ER-dependent secretory pathway thanks to a signal peptide and the second table listed proteins that are predicted to be secreted by non-classical secretory pathways.

1.4 Protein-Protein interactions. The number of PPI identified by querying 3 major (UniProt, BioGrid, and IntAct preselected in the basic analysis) or 9 widely used PPI databases

(BioGrid, IntAct, MINT, MatrixDB, String, Reactome, InnateDB, I2D, and UniProt, chosen thanks to the settings of the custom analysis) are summarized in Table 4. PPI were searched between proteins within (core network) and outside (extended network) the dataset. As expected, the number of PPI identified was higher when 9 rather than 3 databases were queried both for core and extended networks (except sheep and goat, Table 4). Martha and co-workers [46] have previously emphasized the need for a trade-off between the coverage and the reliability of PPI. The maximal coverage can be achieved as we have done by using a traditional union approach for PPI integration, which might however increase false positive PPI. We attempted to minimize the number of false positive by selecting PPI that have been experimentally detected only. With this caution, ProteINSIDE constructs networks with PPI from multiple public PPI databases in order to increase the recovery of species interactome: better informed network with more curated PPI. It is noteworthy that none PPI were identified between Ovine or Caprine ID, which is consistent with the lack of experiments that search for PPI in these species. However, ProteINSIDE database is monthly updated and latest experimentally proven PPI from these species would be identified in the future. Because PPI in ruminant remains poorly recorded, ProteINSIDE automatically constructs networks from ruminant ID within well-studied species chosen either in a blind way by the user or from results of alignments (Blastp) between the sequences of uploaded proteins and their orthologs. It may be useful for biologists working with ruminant species to identify the relationships among different proteins in order to understand the contribution of genes or proteins to pathways or cellular processes.

On the “Protein Interactions” web page of ProteINSIDE, PPI are viewed as tables of listed PPI and as dynamic networks. A dynamic network (Fig 3) links PPI depending on the detection method, the number of interactions by node, the high-centrality (closeness centrality) and the high-degree (betweenness centrality) nodes. To our knowledge, ProteINSIDE is original by the friendly view of networks thanks to the sorting options that provide opportunity to users to identify key nodes (with high closeness and betweenness centralities) both on core and extended networks.

The bench test had provided data in favour of the reliability and accuracy of results produced by ProteINSIDE. We have then tested the ability of ProteINSIDE to produce new hypotheses of research and knowledge for biologists.

2 Analysis of proteomics data from bovine adipose and muscle tissues using ProteINSIDE

Custom analyses were proceeded with 143 and 120 proteins ID identified by proteomics from perirenal AT [4] and *Semitenidosus* muscle [3,5] sampled from the same bovine foetuses at key developmental ages. With these datasets, we aimed firstly to check for the relevance of ProteINSIDE results and secondly to focus on new knowledge provided by ProteINSIDE by comparison with our previous analyses [3–5]. Settings used for custom analyses were: Gene Ontology with IEA, prediction of secreted proteins with default settings, and PPI research on 9 widely used databases (BioGrid, IntAct, MINT, MatrixDB, STRING, Reactome, InnateDB-IMEx, UniProt, and I2D-IMEx).

2.1 Foetal Bovine AT features. ProteINSIDE successfully uploaded and provided a fast overview of the biological information available in UniProt and NCBI databases for 143 proteins that were identified in the AT at each foetal age.

To check for the relevance of ProteINSIDE results, we compared GO terms recovered by ProteINSIDE to those previously published after a GO analysis with DAVID [4]. We reported here enriched GO terms related to biological processes for proteins from clusters 1 (9 proteins highly abundant at 180 dpc [4]; S2 Table) and 8 (20 proteins highly abundant at 260 dpc [4];

[S2 Table](#)). These clusters were chosen for their very different expected GO annotations related to early and late steps of adipogenesis. In agreement with previous GO annotations [4], enriched GO terms recovered by ProteINSIDE were related to cytoskeleton organization, regulation of actin filament depolymerisation, and regulation of apoptotic process for proteins from cluster 1. GO terms recovered for cluster 8 were related to glycolytic process, tricarboxylic acid cycle, and oxidation-reduction processes. In addition for proteins from cluster 8, ProteINSIDE had identified new enriched GO terms such as lipid metabolic process (related to proteins THEM4, ECH1, ACAT2, ECHS1, GPD1, and PCCB) and fatty acid metabolic process (ECHS1, ECH1, PCCB, and THEM4) that are consistent with the highest adipocyte volume and total lipogenic activities in perirenal AT at 260 dpc [4]. These new and relevant annotations recovered by ProteINSIDE may result from a better annotation of bovine proteins taken into account by ProteINSIDE thanks to its regular update.

ProteINSIDE also brought new knowledge since it had predicted 18 proteins as secreted thanks to a signal peptide. Among them, 13 were also annotated by GO terms relative to secretion processes and predicted to be located out of the cell either by TargetP or the subcellular location provided by UniProt ([Table 5](#)). In agreement with our proposed secretome of foetal AT, SERPINA1, APOA1, APOA2, TTR, ALB, TF, HSP90B1, PDIA3, and ADIPOQ were identified in the proteome of bovine plasma [47]; AFP was assayed in the plasma of baboon foetuses [48]. The 3 remaining proteins were reported to be either located on plasma membrane or extracellular vesicular exosome (a membrane-bounded vesicle that is released into the extracellular region) for ERLIN2, in blood for FG, or component of the extracellular matrix for COL6A2 [UniProt Sources]. The 5 proteins (NDUFS8, NDUFS3, NDUFA10, PCCB, and RCN1) predicted to be secreted but not confirmed by GO terms and TargetP were proven to be intracellular and located

Table 5. Potentially secreted proteins with a signal peptide in bovine foetal AT, muscle or both tissues.

Adipose tissue	Muscle tissue	Adipose and muscle tissues
ADIPOQ ^{1,2,3}	ADIPOQ ^{1,2,3}	ADIPOQ
NDUS3	ALB ^{1,2,3}	APOA1
ERLIN2 ^{2,3}	AFP ^{2,3}	AFP
NDUS8	APOA1 ^{1,2,3}	ALB
NDUFA10	GSN ^{1,2,3}	PDIA3
SERPINA1 ^{1,2,3}	GARS ¹	SERPINA1
APOA1 ^{1,2,3}	SERPINA1 ^{1,2,3}	
APOA2 ^{1,2,3}	P4HB ^{1,2,3}	
FGG ^{1,2,3}	PDIA3 ^{1,2}	
TTR ^{1,2,3}		
ALB ^{1,2,3}		
AFP ^{1,3}		
TF ^{1,2,3}		
PCCB		
COL6A2 ^{1,2,3}		
HSP90B1 ^{2,3}		
PDIA3 ^{2,3}		
RCN1		

¹ confirmed by GO terms

² confirmed by TargetP

³ confirmed by Subcellular location provided by UniProt resource

doi:10.1371/journal.pone.0128086.t005

in the mitochondrion or endoplasmic reticulum [UniProt Sources]. They were listed as secreted because SignalP identified a signal peptide on their amino acid sequences for an intracellular rather than plasma membrane translocation (false positive). These results highlight the usefulness of the integration of results from both prediction and annotations for a reliable list of proteins proposed to be secreted. ProteINSIDE had also predicted 89 potentially secreted proteins without signal peptide. However, only 11 proteins were confirmed by GO terms, TargetP or known subcellular location in UniProt (PRDX6, MDH1, ALAD, VCP, CAPZA1, CAPZA2, TPI1, ARHGDIA, HNRNPK, EIF5A, and TPMT). Among them, 6 proteins are known to be secreted by AT (PDRX6 [49]; MDH1 and TPI1 [50]; VCP, ARHGDIA, and EIF5A [51]). Thus, by using ProteINSIDE, we have identified 13 proteins potentially secreted thanks to a signal peptide and 11 proteins potentially secreted by non-classical ways. It is noteworthy that these proteins could be synthesized by either AT such as adiponectin [4] or other tissues and taken up by AT, that remains to be questioned by the search of their mRNA in AT in available published results or by new experiments.

The PPI module of ProteINSIDE had revealed that 107 proteins of the dataset were linked by 300 interactions (Fig 5). The synthetic view of PPI as network had shown 5 subnetworks of proteins contributing to a same process: cell proliferation (Fig 5D), proteasome complex (Fig

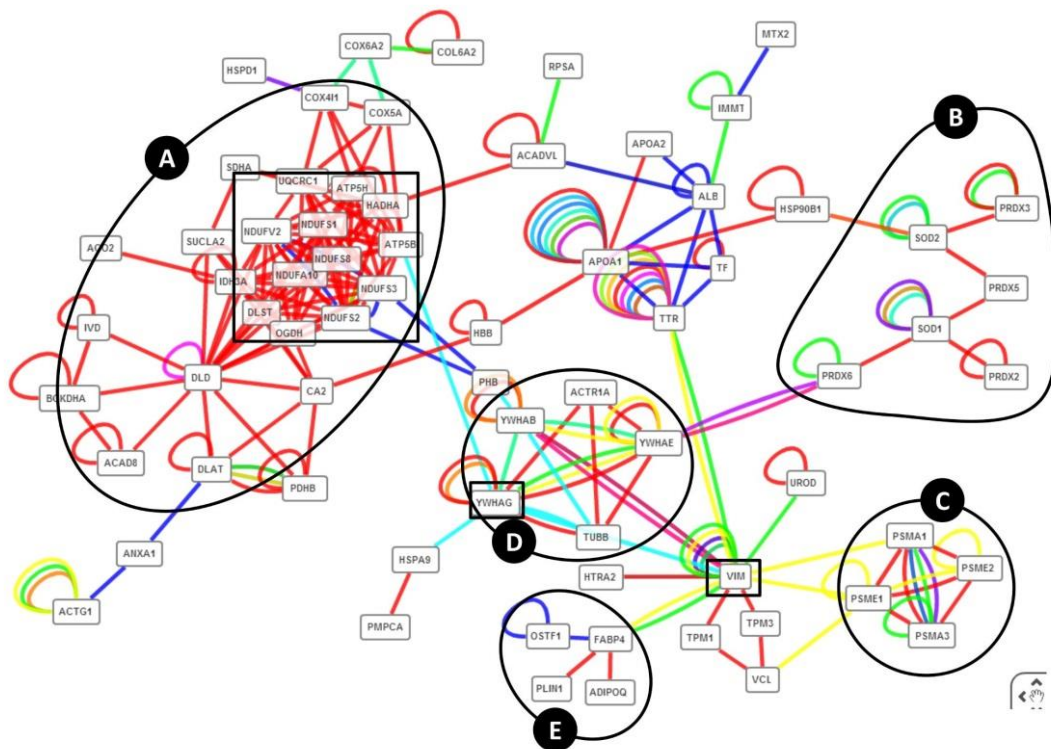


Fig 5. Interactions network between proteins of the bovine foetal AT. We highlighted the relationships among different proteins involved in a same process: (A) mitochondrial metabolism, (B) redox activity, (C) proteasome complex, (D) cell proliferation, and (E) differentiation and metabolism of AT. Squares indicate key proteins identified by sorting the network with algorithms of betweenness (an average of 400) and closeness centralities (an average of 0.3).

doi:10.1371/journal.pone.0128086.g005

5C), complex I and III of the respiratory chain (Fig 5A), redox activity (Fig 5B), and differentiation and metabolism of AT (Fig 5E). Then, to identify key proteins in the foetal AT growth we have sorted the network by applying high values of betweenness and closeness centralities and we proposed 15 proteins as centrals in the network (Fig 5). Among them, 13 were mitochondrial proteins involved in Krebs cycle, respiratory chain, and branched chain amino acid metabolism, in agreement with metabolic pathways that were proposed to control the increase in foetal adipocyte size in our previous analysis [4]. The last 2 key proteins were YWHAG previously proposed to balance cell cycle progression and cell cycle arrest that are two major cellular events for adipocyte precursor proliferation and differentiation, as well as Vimentin (VIM) that was not identified as key protein in our previous data mining [4]. However, VIM was reported to be over-abundant at 110 and 180 dpc or with a stable abundance across foetal ages depending on the isoforms. VIM, a class-III intermediate filament found in various non-epithelial cells, especially mesenchymal cells was first proposed to sustain the well-described dramatic change in cell shape from fusiform adipocyte precursors to spherical adipocytes [4] during adipogenesis. More recent results have shown that VIM contributes to both the formation of lipid droplets through an interaction with perilipin [52] and the regulation of fatty acid transport through an interaction with FABP4 [53]. Thus, the central role of VIM in the foetal growth of AT proposed by ProteINSIDE is sustained by experimental results that emphasized its role in early and late stages of adipogenesis.

2.2 Foetal Bovine muscle tissue features. ProteINSIDE correctly identified 120 muscular proteins present at all the five foetal ages [3,5] and provided a fast overview of biological information available in UniProt and NCBI databases.

As for data from AT, the relevance of ProteINSIDE results was checked. We have compared enriched GO terms recovered by ProteINSIDE to those previously published [3]. We reported here enriched GO terms related to biological processes for proteins from cluster 8 (18 proteins over-abundant at 60 dpc [3,5]; S2 Table) and cluster 1 (14 proteins over-abundant at 180 dpc [3,5]; S2 Table). In agreement with previous GO annotations [3,5], enriched GO terms recovered by ProteINSIDE were related to the negative regulation of apoptotic process, DNA binding, and skeletal muscle thin filament assembly for cluster 8, and to glycolytic process and the regulation of translational initiation for cluster 1. These GO related to cellular and molecular events reflect the proliferation of muscular cells before 180 dpc and, their metabolic differentiations, after in accordance with the know pivotal role of the foetal age 180 dpc [6].

As new knowledge, ProteINSIDE had predicted 9 secreted proteins among the 120 proteins from foetal muscle (Table 5). Among them, 7 proteins were confirmed by TargetP, subcellular location in UniProt, and GO terms related to secretory pathways while PDIA3 and GARS were only confirmed by one GO term. Available data from proteomics or specific isolations in bovine [47], human [54,55] and baboon [48] plasma confirm that APOA1, P4HB, GSN, ADIPOQ, AFP, ALB, PDIA3, and SERPINA1 are secreted proteins. The protein GARS was a false positive of the ProteINSIDE prediction since available experimental data recorded in UniProt have located this protein in the cytosol or the mitochondrion [UniProt sources: <http://www.uniprot.org/uniprot/P41250>]. ProteINSIDE had also predicted 73 proteins secreted by cellular pathways that do not involved signal peptide. However, 11 proteins (F3, CAPZA2, TPI1, ARHGDI1, HNRNP1, EIF5A, SNX6, ANXA2, COPS4, PRDX6, and AKR1B10) were confirmed by TargetP, GO terms or subcellular location in UniProt. Among them, 6 proteins are known to be secreted by skeletal muscle (TPI1, ARHGDI1, HNRNP1, EIF5A, PRDX6 [56], and ANXA2 [57]). Thus, with ProteINSIDE we have identified 8 and 11 proteins potentially secreted by a signal peptide or by other secretory pathways, respectively. As for AT, the remaining question is whether these proteins are synthesised or taken up from the blood by the muscle. Available results from mRNA tissue-distribution show that AFP is synthesized by the

liver and taken up by the muscle, since no mRNA was detected in this tissue [48]. The synthesis of ADIPOQ (originally thought to be limited to adipocytes) by muscular cells remains controversial because of the very low mRNA abundance in L6 myotubes when compared to white AT [58] and, its immunohistochemical location in vascular endothelium and adipocytes of the gastrocnemius muscle [59].

ProteINSIDE revealed 171 interactions between 76 proteins of the dataset (Fig 6). A synthetic view of PPI as network had shown 4 subnetworks of proteins contributing to a same process: the muscle development (Fig 6A), cell proliferation (Fig 6D), energetic complex (Fig 6B), and respiratory chain (Fig 6C) in agreement with the previous pathways proposed to sustain muscle growth in bovine foetuses [3,5]. Using high levels of betweenness and closeness centralities, we proposed for the first time 5 proteins as central in this network and for foetal muscle growth (square in Fig 6): YWHAZ and YWHAE regulating the balance between cell cycle progression and apoptosis, MYC an activator of the transcription of growth-related genes, and HSPB1 and HSPA8 involved in the protection of structural proteins (like desmin, actin, and titin) and also known as anti-apoptotic factors, promoting cell survival. Interestingly, HSPB1 was identified as a central protein in a network of protein related to muscle hypertrophy and functioning in growing bovine [60]. Moreover, HSPB1 was proposed to regulate the total number of fibres in Bovine foetuses (B. Picard, unpublished data). Present results suggest that HSPB1 may also contribute to myogenesis that is sustained by a very recent study reporting a

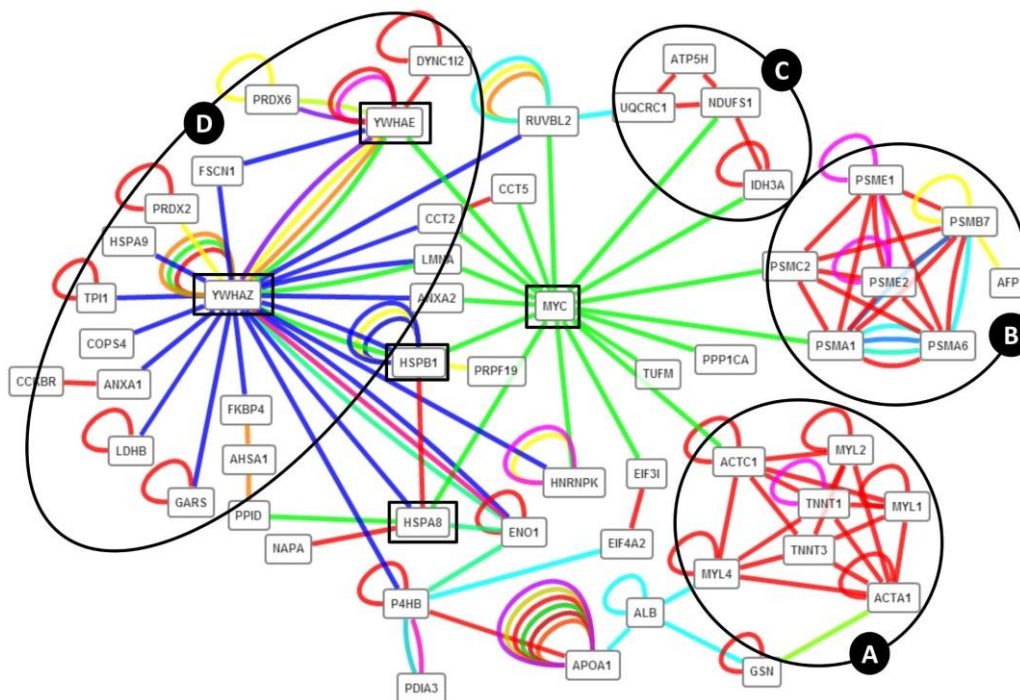


Fig 6. PPI network between proteins from the bovine foetal muscle. We highlighted the relationships among different proteins involved in a same process: (A) muscle development, (B) energetic complexes, (C) respiratory chain, and (D) cell proliferation. Squares indicate key proteins identified by sorting the network with algorithms of betweenness (an average of 100) and closeness centralities (an average of 0.4).

doi:10.1371/journal.pone.0128086.g006

repressed formation of myotubes by siRNA inhibition of HSPB1 expression in bovine myogenic cell, while expression of HSPB1 enhanced the expression of desmin and accelerated formation of myotubes [61].

2.3 An union analysis of adipose and muscular data to identify pathways involved in the growth of AT and muscle and their cross-talk. There is striking evidence for developmental and functional links between muscle and AT: the successive waves of growth of muscle and AT suggest a priority for muscle growth, and the comparison between lean and fat bovine breeds suggest that an increased muscular development is concomitant with a decrease in AT mass [2]. In order to increase our knowledge on pathways involved in the growth of both AT and muscle and in the cross-talk between them, we have compared the adipose and muscular datasets. We have identified 46 proteins present both in adipose and muscle tissues that were subjected to a custom analysis with the following settings: GO annotation without IEA codes, prediction of secreted proteins, PPI research on 9 widely used databases (BioGrid, IntAct, MINT, MatrixDB, STRING, Reactome, InnateDB-IMEx, UniProt, and I2D-IMEx), and an extended network of PPI with proteins outside of the dataset in Human.

The 46 proteins were annotated by enriched GO terms (p -value < 0.001 corrected by the BH test) related to apoptosis (negative regulation of apoptotic process, apoptotic process, and regulation of apoptotic process), the progression of cell cycle (DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle, anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process, and G1/S transition of mitotic cell cycle) metabolic processes (regulation of cellular amino acid metabolic process, respiratory electron transport chain, glucose metabolic process, and protein polyubiquitination), the regulation of cytokine- or interleukin-mediated signaling pathways (negative regulation of tumor necrosis factor-mediated signalling pathway and negative regulation of interleukin-8 secretion). Six of these 46 proteins were predicted and confirmed by GO terms and TargetP as being secreted because of a signal peptide (Table 5). As previously discussed in sections 2.1 and 2.2, all of them were reported in the bovine plasma [48]. The identification of ADIPOQ in both bovine foetal tissues, and especially in the muscle, is surprising because intramuscular adipocytes believed to synthesize adiponectin [59], were histologically characterised after birth in bovine [2]. Thus, to question the synthesis and secretion or the uptake of adiponectin by the foetal muscle we assayed its mRNA abundance (Fig 7). The very low abundance of ADIPOQ mRNA in foetal muscle, when compared to adult or foetal AT, suggests that ADIPOQ is rather taken up than synthesized by foetal muscle. The role of the adipose ADIPOQ [4] within muscle remained to be studied during the bovine foetal growth. ProteINSIDE had also predicted 32 potentially secreted proteins without signal peptide. However, only 5 proteins were both confirmed by GO terms and TargetP (CAPZA2, TPI1, ARHGDIA, HNRNPK, and EIF5A), and a last protein was also confirmed by Subcellular location (PRD6X). As described in section 2.1 and 2.2 some of these proteins are known to be secreted by AT and skeletal muscle in adult (PRD6X, TPI1, ARHGDIA, and EIF5A).

Among the 46 common proteins, 23 were linked by 40 interactions and grouped in 2 sub-networks of proteins related to muscle proteasome and mitochondrial metabolism (already described in Fig 5C and 5A, respectively). In order to identify new putative key proteins for the foetal growth of AT and muscle, we have extended the network by including proteins known to interact with the 46 proteins in Human. The extended network linked 1063 proteins (41 from the dataset) by 1400 interactions. By using high values of betweenness and closeness centralities, we proposed 35 proteins (18 from dataset; Fig 8) as being central in the extended network. Among them, 6 proteins (external of the dataset) were highly central: VCAM1, GABARAPL1, GABARAPL2, GABARAP, MAP1LC3A, and MAP1LC3B. The presence of these proteins or

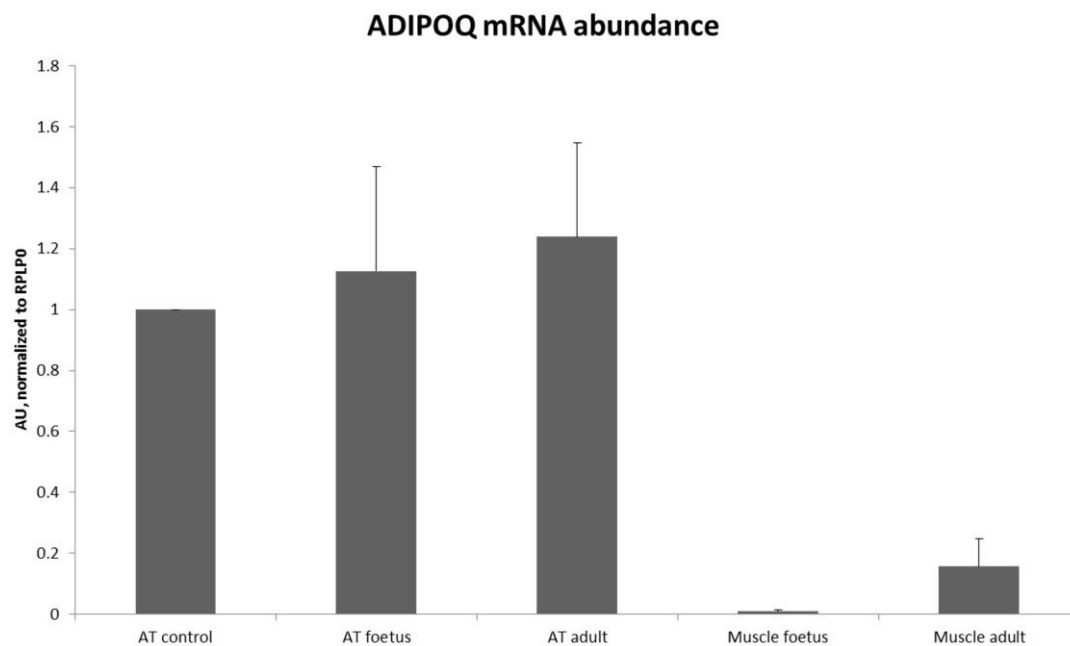


Fig 7. mRNA abundance of adiponectin (ADIPOQ) in bovine AT and muscle. The abundance of ADIPOQ was normalized to the mRNA abundance of ribosomal protein P0 (RPLP0). Results are $\Delta\Delta$ CT for foetal and adult AT or muscle samples, relatively to a control sample that is an adult AT. PCR were carried out as previously described [62].

doi:10.1371/journal.pone.0128086.g007

their mRNA in adipose and muscular cells was reported for VCAM1 [63,64], GABARAPL1 [65,66], GABARAPL2 [64,67], GABARAP [64,68], MAP1LC3A [64,69], and MAP1LC3B [64,70].

VCAM1 also named CD106 is a cell surface marker of mesenchymal stem cells found in AT [71] and muscle [72]. The presence of VCAM1 in foetal bovine tissue remains to be shown but should indicate that mesenchymal stem cells contribute to the post-embryonic growth of AT and muscle as already proposed [2]. GABARAPL1, GABARAPL2, GABARAP, MAP1LC3A, and MAP1LC3B are members of Atg8 family in higher eukaryotes and involved in autophagy, a major catabolic pathway that impacts several cellular pathways such as cell survival, differentiation, tumorigenesis. . . The centrality of these autophagy-related proteins in the extended network suggests that autophagy plays a role in the differentiation and growth of adipose and muscle tissues in bovine foetuses. This remains to be studied. However, this hypothesis is supported by recent *in vitro* data that demonstrate that autophagy is increased and required for the differentiation of C2C12 myoblasts [73] and 3T3-L1 preadipocytes [74] probably to sustain the significant morphological and biochemical remodelling of differentiating cells. Moreover *in vivo*, the deletion of an autophagy-related gene in mice impairs foetal adipogenesis, consequently pups have only 15% of differentiated adipocytes in subcutaneous AT when compared to wild-type counterparts [75]. The deletion of an autophagy-related gene in Myf5+ progenitor cells impairs brown AT differentiation and function, muscle differentiation, reduces muscle mass, and leads to glucose intolerance in mice [76]. The study of the expression of autophagy-related genes by comparison with some markers of adipogenesis and myogenesis or secreted

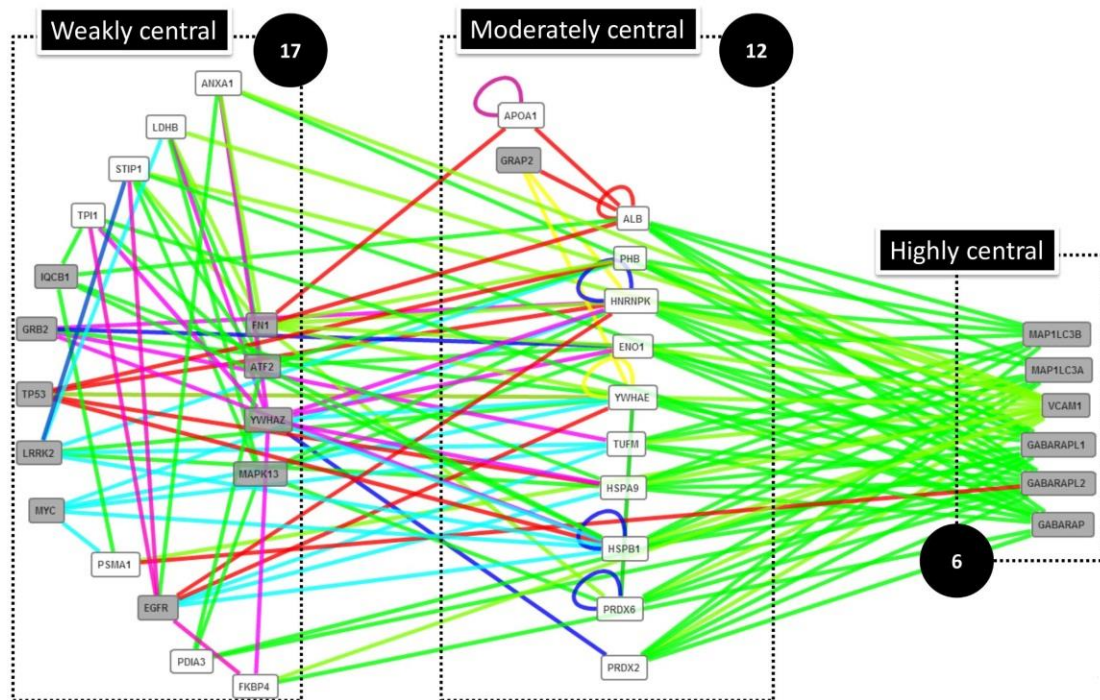


Fig 8. Extended PPI network that includes proteins identified both in foetal AT and muscle. ProteINSIDE built a network between the 46 proteins identified in AT and muscle, and proteins outside of the dataset and know to interact with them in Human species. We filtered and sorted the network using high values of betweenness (an average of 10000) and closeness centralities (an average of 0,318). We have identified 6 proteins that were highly central in the dataset (linked with the maximum of proteins and pathways), 12 proteins were moderately central (engaged in the maximum of pathways but not necessary with many proteins), and 17 proteins were more weakly central (less linked with the maximum of proteins and less engaged in pathways, but central on the network). White boxes indicate proteins that are from the bovine dataset and grey boxes indicate proteins that are external to the dataset.

doi:10.1371/journal.pone.0128086.g008

factors that we have identified, such as adiponectin, should help to understand the growth of AT relatively to muscle in bovine foetuses.

Conclusions

Using a list of genes or proteins, ProteINSIDE proceeds to a fast overview of biological information for ID of the uploaded dataset, a functional annotation according to GO, a prediction of secreted proteins, and a simple and interactive view of known and experimentally proven PPI. The similar or better reliability of results produced by ProteINSIDE was demonstrated by a bench test involving others available resources. Moreover, we have verified the biological relevance of the results by comparison with our previous analyses. Lastly, we have used ProteINSIDE to propose new hypotheses of research that should help to better understand the growth of AT and muscle in bovine. Among the new hypotheses of research that deserved to be investigated, we have focused on an uptake of adiponectin by the foetal muscle and the role of autophagy on the ontogenesis of AT and muscle. To go further, ProteINSIDE can easily be upgraded to broaden the range of species (as other farm species: e.g. pig, chicken. . .) by an update of the database and the workflow.

Supporting Information

S1 Table. Lists of 1000 proteins ID from the 6 species used to test ProteINSIDE's performances.

(XLSX)

S2 Table. Lists adipose and muscle tissues proteins from Taga et al. [4], and Chaze et al. [3,5].

(XLSX)

Acknowledgments

The authors acknowledge A. Delavaud for conducting the mRNA quantification.

Author Contributions

Conceived and designed the experiments: NK MB. Analyzed the data: NK MB. Contributed reagents/materials/analysis tools: NK JT MR MB. Wrote the paper: NK MB BP.

References

1. Woelders H, Te Pas MF, Bannink A, Veerkamp RF, Smits MA (2011) Systems biology in animal sciences. *Animal* 5: 1036–1047. doi: [10.1017/S1751731111000036](https://doi.org/10.1017/S1751731111000036) PMID: [22440099](https://pubmed.ncbi.nlm.nih.gov/22440099/)
2. 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: 1093–1109. doi: [10.1017/S1751731110000601](https://doi.org/10.1017/S1751731110000601) PMID: [22444612](https://pubmed.ncbi.nlm.nih.gov/22444612/)
3. Chaze T, Meunier B, Chambon C, Jurie C, Picard B (2008) In vivo proteome dynamics during early bovine myogenesis. *Proteomics* 8: 4236–4248. doi: [10.1002/pmic.200701101](https://doi.org/10.1002/pmic.200701101) PMID: [18924180](https://pubmed.ncbi.nlm.nih.gov/18924180/)
4. Taga H, Chilliard Y, Meunier B, Chambon C, Picard B, Zingaretti MC, 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. doi: [10.1002/jcp.22893](https://doi.org/10.1002/jcp.22893) PMID: [21678425](https://pubmed.ncbi.nlm.nih.gov/21678425/)
5. Chaze T, Meunier B, Chambon C, Jurie C, Picard B (2009) Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle. *Animal* 3: 980–1000. doi: [10.1017/S1751731109004315](https://doi.org/10.1017/S1751731109004315) PMID: [22444818](https://pubmed.ncbi.nlm.nih.gov/22444818/)
6. Picard B, Cassar-Malek I, Guillemin N, Bonnet M (2011) Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: Moo-Young M, editor. *Comprehensive Biotechnology* (Second Edition). Burlington: Academic Press. pp. 395–405.
7. Hudson NJ, Lyons RE, Reverter A, Greenwood PL, Dalrymple BP (2013) Inferring the in vivo cellular program of developing bovine skeletal muscle from expression data. *Gene Expr Patterns* 13: 109–125. doi: [10.1016/j.gep.2013.02.001](https://doi.org/10.1016/j.gep.2013.02.001) PMID: [23419240](https://pubmed.ncbi.nlm.nih.gov/23419240/)
8. Bionaz M, Loor JJ (2012) Ruminant metabolic systems biology: reconstruction and integration of transcriptome dynamics underlying functional responses of tissues to nutrition and physiological state. *Gene Regul Syst Bio* 6: 109–125. doi: [10.4137/GRSB.S9852](https://doi.org/10.4137/GRSB.S9852) PMID: [22807626](https://pubmed.ncbi.nlm.nih.gov/22807626/)
9. Shahzad K, Loor JJ (2012) Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism. *Curr Genomics* 13: 379–394. doi: [10.2174/138920212801619269](https://doi.org/10.2174/138920212801619269) PMID: [23372424](https://pubmed.ncbi.nlm.nih.gov/23372424/)
10. Romao JM, Jin W, He M, McAllister T, Guan le L (2013) Elucidation of molecular mechanisms of physiological variations between bovine subcutaneous and visceral fat depots under different nutritional regimes. *PLoS One* 8: e83211. doi: [10.1371/journal.pone.0083211](https://doi.org/10.1371/journal.pone.0083211) PMID: [24349463](https://pubmed.ncbi.nlm.nih.gov/24349463/)
11. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25: 288–289. doi: [10.1093/bioinformatics/btn615](https://doi.org/10.1093/bioinformatics/btn615) PMID: [19033274](https://pubmed.ncbi.nlm.nih.gov/19033274/)
12. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48. doi: [10.1186/1471-2105-10-48](https://doi.org/10.1186/1471-2105-10-48) PMID: [19192299](https://pubmed.ncbi.nlm.nih.gov/19192299/)
13. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25: 3045–3046. doi: [10.1093/bioinformatics/btp536](https://doi.org/10.1093/bioinformatics/btp536) PMID: [19744993](https://pubmed.ncbi.nlm.nih.gov/19744993/)

14. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. (2010) GeneCards Version 3: the human gene integrator. Database (Oxford) 2010: baq020. doi: [10.1093/database/baq020](https://doi.org/10.1093/database/baq020) PMID: [20689021](https://pubmed.ncbi.nlm.nih.gov/20689021/)
15. Tarcea VG, Weymouth T, Ade A, Bookvich A, Gao J, Mahavisno V, et al. (2009) Michigan molecular interactions r2: from interacting proteins to pathways. Nucleic Acids Res 37: D642–646. doi: [10.1093/nar/gkn722](https://doi.org/10.1093/nar/gkn722) PMID: [18978014](https://pubmed.ncbi.nlm.nih.gov/18978014/)
16. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. (2012) The IntAct molecular interaction database in 2012. Nucleic Acids Res 40: D841–846. doi: [10.1093/nar/gkr1088](https://doi.org/10.1093/nar/gkr1088) PMID: [22121220](https://pubmed.ncbi.nlm.nih.gov/22121220/)
17. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. (2013) The BioGRID interaction database: 2013 update. Nucleic Acids Res 41: D816–823. doi: [10.1093/nar/gks1158](https://doi.org/10.1093/nar/gks1158) PMID: [23203989](https://pubmed.ncbi.nlm.nih.gov/23203989/)
18. Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautaud E, et al. (2011) PSICQUIC and PSISCOPE: accessing and scoring molecular interactions. Nat Methods 8: 528–529. doi: [10.1038/nmeth.1637](https://doi.org/10.1038/nmeth.1637) PMID: [21716279](https://pubmed.ncbi.nlm.nih.gov/21716279/)
19. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41: D808–815. doi: [10.1093/nar/gks1094](https://doi.org/10.1093/nar/gks1094) PMID: [23203871](https://pubmed.ncbi.nlm.nih.gov/23203871/)
20. Hernandez-Toro J, Prieto C, De las Rivas J (2007) APID2NET: unified interactome graphic analyzer. Bioinformatics 23: 2495–2497. PMID: [17644818](https://pubmed.ncbi.nlm.nih.gov/17644818/)
21. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277. PMID: [10827456](https://pubmed.ncbi.nlm.nih.gov/10827456/)
22. Frank K, Sippl MJ (2008) High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics 24: 2172–2176. doi: [10.1093/bioinformatics/btn422](https://doi.org/10.1093/bioinformatics/btn422) PMID: [18697773](https://pubmed.ncbi.nlm.nih.gov/18697773/)
23. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785–786. doi: [10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701) PMID: [21959131](https://pubmed.ncbi.nlm.nih.gov/21959131/)
24. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57. doi: [10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211) PMID: [19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
25. Ramirez F, Lawyer G, Albrecht M (2012) Novel search method for the discovery of functional relationships. Bioinformatics 28: 269–276. doi: [10.1093/bioinformatics/btr631](https://doi.org/10.1093/bioinformatics/btr631) PMID: [22180409](https://pubmed.ncbi.nlm.nih.gov/22180409/)
26. Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37: W305–311. doi: [10.1093/nar/gkp427](https://doi.org/10.1093/nar/gkp427) PMID: [19465376](https://pubmed.ncbi.nlm.nih.gov/19465376/)
27. Ulitsky I, Maron-Katz A, Shavit S, Sagir D, Linhart C, Elkorn R, et al. (2010) Expander: from expression microarrays to networks and functions. Nat Protoc 5: 303–322. doi: [10.1038/nprot.2009.230](https://doi.org/10.1038/nprot.2009.230) PMID: [20134430](https://pubmed.ncbi.nlm.nih.gov/20134430/)
28. Nikitin A, Egorov S, Daraselia N, Mazo I (2003) Pathway studio—the analysis and navigation of molecular networks. Bioinformatics 19: 2155–2157. PMID: [14594725](https://pubmed.ncbi.nlm.nih.gov/14594725/)
29. McCarthy FM, Gresham CR, Buza TJ, Chouvarine P, Pillai LR, Kumar R, et al. (2011) AgBase: supporting functional modeling in agricultural organisms. Nucleic Acids Res 39: D497–506. doi: [10.1093/nar/gkq1115](https://doi.org/10.1093/nar/gkq1115) PMID: [21075795](https://pubmed.ncbi.nlm.nih.gov/21075795/)
30. Caccia D, Dugo M, Callari M, Bongarzone I (2013) Bioinformatics tools for secretome analysis. Biochim Biophys Acta 1834: 2442–2453. doi: [10.1016/j.bbapap.2013.01.039](https://doi.org/10.1016/j.bbapap.2013.01.039) PMID: [23395702](https://pubmed.ncbi.nlm.nih.gov/23395702/)
31. Romacho T, Elsen M, Rohrborn D, Eckel J (2014) Adipose tissue and its role in organ crosstalk. Acta Physiol (Oxf) 210: 733–753. doi: [10.1111/apha.12246](https://doi.org/10.1111/apha.12246) PMID: [24495317](https://pubmed.ncbi.nlm.nih.gov/24495317/)
32. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011: bar009.
33. Coordinators NR (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41: D8–D20. doi: [10.1093/nar/gks1189](https://doi.org/10.1093/nar/gks1189) PMID: [23193264](https://pubmed.ncbi.nlm.nih.gov/23193264/)
34. Al-Shahrour F, Diaz-Uriarte R, Dopazo J (2004) FatGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 20: 578–580. PMID: [14990455](https://pubmed.ncbi.nlm.nih.gov/14990455/)
35. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289–300.
36. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29. PMID: [10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)

37. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26: 2347–2348. doi: [10.1093/bioinformatics/btq430](https://doi.org/10.1093/bioinformatics/btq430) PMID: [20656902](https://pubmed.ncbi.nlm.nih.gov/20656902/)
38. Nickel W (2003) The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *Eur J Biochem* 270: 2109–2119. PMID: [12752430](https://pubmed.ncbi.nlm.nih.gov/12752430/)
39. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953–971. PMID: [17446895](https://pubmed.ncbi.nlm.nih.gov/17446895/)
40. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005–1016. PMID: [10891285](https://pubmed.ncbi.nlm.nih.gov/10891285/)
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
42. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D (2008) A protein interaction network associated with asthma. *J Theor Biol* 252: 722–731. doi: [10.1016/j.jtbi.2008.02.011](https://doi.org/10.1016/j.jtbi.2008.02.011) PMID: [18395227](https://pubmed.ncbi.nlm.nih.gov/18395227/)
43. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigan C, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* 12: 1611–1618. PMID: [12368254](https://pubmed.ncbi.nlm.nih.gov/12368254/)
44. Hiller K, Grote A, Scheer M, Munch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* 32: W375–379. PMID: [15215414](https://pubmed.ncbi.nlm.nih.gov/15215414/)
45. Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35: W429–432. PMID: [17483518](https://pubmed.ncbi.nlm.nih.gov/17483518/)
46. Martha VS, Liu Z, Guo L, Su Z, Ye Y, Fang H, et al. (2011) Constructing a robust protein-protein interaction network by integrating multiple public databases. *BMC Bioinformatics* 12 Suppl 10: S7. doi: [10.1186/1471-2105-12-S10-S7](https://doi.org/10.1186/1471-2105-12-S10-S7) PMID: [22165958](https://pubmed.ncbi.nlm.nih.gov/22165958/)
47. Henning AK, Groschup MH, Mettenleiter TC, Karger A (2014) Analysis of the bovine plasma proteome by matrix-assisted laser desorption/ionisation time-of-flight tandem mass spectrometry. *Vet J* 199: 175–180. doi: [10.1016/j.tvjl.2013.10.029](https://doi.org/10.1016/j.tvjl.2013.10.029) PMID: [24268478](https://pubmed.ncbi.nlm.nih.gov/24268478/)
48. Naval J, Calvo M, Laborda J, Dubouch P, Frain M, Sala-Trepal JM, et al. (1992) Expression of mRNAs for alpha-fetoprotein (AFP) and albumin and incorporation of AFP and docosahexaenoic acid in baboon fetuses. *J Biochem* 111: 649–654. PMID: [1379224](https://pubmed.ncbi.nlm.nih.gov/1379224/)
49. Chiellini C, Cochet O, Negroni L, Samson M, Poggi M, Ailhaud G, et al. (2008) Characterization of human mesenchymal stem cell secretome at early steps of adipocyte and osteoblast differentiation. *BMC Mol Biol* 9: 26. doi: [10.1186/1471-2199-9-26](https://doi.org/10.1186/1471-2199-9-26) PMID: [18302751](https://pubmed.ncbi.nlm.nih.gov/18302751/)
50. Roelofsen H, Dijkstra M, Weening D, de Vries MP, Hoek A, Vonk RJ (2009) Comparison of isotope-labeled amino acid incorporation rates (CILAIR) provides a quantitative method to study tissue secretomes. *Mol Cell Proteomics* 8: 316–324. doi: [10.1074/mcp.M800254-MCP200](https://doi.org/10.1074/mcp.M800254-MCP200) PMID: [18840871](https://pubmed.ncbi.nlm.nih.gov/18840871/)
51. Zhong J, Krawczyk SA, Chaerkady R, Huang H, Goel R, Bader JS, et al. (2010) Temporal profiling of the secretome during adipogenesis in humans. *J Proteome Res* 9: 5228–5238. doi: [10.1021/pr100521c](https://doi.org/10.1021/pr100521c) PMID: [20707391](https://pubmed.ncbi.nlm.nih.gov/20707391/)
52. Heid H, Rickelt S, Zimbelmann R, Winter S, Schumacher H, Dorflinger Y, et al. (2014) On the formation of lipid droplets in human adipocytes: the organization of the perilipin-vimentin cortex. *PLoS One* 9: e90386. doi: [10.1371/journal.pone.0090386](https://doi.org/10.1371/journal.pone.0090386) PMID: [24587346](https://pubmed.ncbi.nlm.nih.gov/24587346/)
53. Bag S, Ramaiah S, Anbarasu A (2014) *fabp4* is central to eight obesity associated genes: A functional gene network-based polymorphic study. *J Theor Biol*.
54. Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, et al. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics* 3: 311–326. PMID: [14718574](https://pubmed.ncbi.nlm.nih.gov/14718574/)
55. Kwiatkowski DJ, Stossel TP, Orkin SH, Mole JE, Colten HR, Yin HL (1986) Plasma and cytoplasmic gelsolins are encoded by a single gene and contain a duplicated actin-binding domain. *Nature* 323: 455–458. PMID: [3020431](https://pubmed.ncbi.nlm.nih.gov/3020431/)
56. Hartwig S, Raschke S, Knebel B, Scheler M, Irmeler M, Passlack W, et al. (2014) Secretome profiling of primary human skeletal muscle cells. *Biochim Biophys Acta* 1844: 1011–1017. doi: [10.1016/j.bbapap.2013.08.004](https://doi.org/10.1016/j.bbapap.2013.08.004) PMID: [23994228](https://pubmed.ncbi.nlm.nih.gov/23994228/)
57. Le Bihan MC, Bigot A, Jensen SS, Dennis JL, Rogowska-Wrzesinska A, Laine J, et al. (2012) In-depth analysis of the secretome identifies three major independent secretory pathways in differentiating human myoblasts. *J Proteomics* 77: 344–356. doi: [10.1016/j.jprot.2012.09.008](https://doi.org/10.1016/j.jprot.2012.09.008) PMID: [23000592](https://pubmed.ncbi.nlm.nih.gov/23000592/)
58. Krause MP, Liu Y, Vu V, Chan L, Xu A, Riddell MC, et al. (2008) Adiponectin is expressed by skeletal muscle fibers and influences muscle phenotype and function. *Am J Physiol Cell Physiol* 295: C203–212. doi: [10.1152/ajpcell.00030.2008](https://doi.org/10.1152/ajpcell.00030.2008) PMID: [18463233](https://pubmed.ncbi.nlm.nih.gov/18463233/)

59. Yang B, Chen L, Qian Y, Triantafyllou JA, McNulty JA, Carrick K, et al. (2006) Changes of skeletal muscle adiponectin content in diet-induced insulin resistant rats. *Biochem Biophys Res Commun* 341: 209–217. PMID: [16414018](#)
60. Guillemain N, Bonnet M, Jurie C, Picard B (2011) Functional analysis of beef tenderness. *J Proteomics* 75: 352–365. doi: [10.1016/j.jprot.2011.07.026](#) PMID: [21855665](#)
61. Zhang Q, Lee HG, Kang SK, Baik M, Choi YJ (2014) Heat-shock protein beta 1 regulates androgen-mediated bovine myogenesis. *Biotechnol Lett* 36: 1225–1231. doi: [10.1007/s10529-014-1489-2](#) PMID: [24563320](#)
62. Bonnet M, Bernard L, Bes S, Leroux C (2013) Selection of reference genes for quantitative real-time PCR normalisation in adipose tissue, muscle, liver and mammary gland from ruminants. *Animal* 7: 1344–1353. doi: [10.1017/S1751731113000475](#) PMID: [23552195](#)
63. Raschke S, Eckardt K, Bjorklund Holven K, Jensen J, Eckel J (2013) Identification and validation of novel contraction-regulated myokines released from primary human skeletal muscle cells. *PLoS One* 8: e62008. doi: [10.1371/journal.pone.0062008](#) PMID: [23637948](#)
64. Scheideler M, Elabd C, Zaragosi LE, Chiellini C, Hackl H, Sanchez-Cabo F, et al. (2008) Comparative transcriptomics of human multipotent stem cells during adipogenesis and osteoblastogenesis. *BMC Genomics* 9: 340. doi: [10.1186/1471-2164-9-340](#) PMID: [18637193](#)
65. Canello R, Henegar C, Viguier N, Taleb S, Poitou C, Rouault C, et al. (2005) Reduction of macrophage infiltration and chemoattractant gene expression changes in white adipose tissue of morbidly obese subjects after surgery-induced weight loss. *Diabetes* 54: 2277–2286. PMID: [16046292](#)
66. He H, Liu X (2013) Characterization of transcriptional complexity during longissimus muscle development in bovines using high-throughput sequencing. *PLoS One* 8: e64356. doi: [10.1371/journal.pone.0064356](#) PMID: [23762238](#)
67. Welle S, Tawil R, Thornton CA (2008) Sex-related differences in gene expression in human skeletal muscle. *PLoS One* 3: e1385. doi: [10.1371/journal.pone.0001385](#) PMID: [18167544](#)
68. Sadkowski T, Ciecierska A, Majewska A, Oprzadek J, Dasiewicz K, Ollik M, et al. (2014) Transcriptional background of beef marbling—novel genes implicated in intramuscular fat deposition. *Meat Sci* 97: 32–41. doi: [10.1016/j.meatsci.2013.12.017](#) PMID: [24491505](#)
69. Welle S, Cardillo A, Zanche M, Tawil R (2009) Skeletal muscle gene expression after myostatin knock-out in mature mice. *Physiol Genomics* 38: 342–350. doi: [10.1152/physiolgenomics.00054.2009](#) PMID: [19509079](#)
70. Yoon JH, Song P, Jang JH, Kim DK, Choi S, Kim J, et al. (2011) Proteomic analysis of tumor necrosis factor-alpha (TNF-alpha)-induced L6 myotube secretome reveals novel TNF-alpha-dependent myokines in diabetic skeletal muscle. *J Proteome Res* 10: 5315–5325. doi: [10.1021/pr200573b](#) PMID: [22023146](#)
71. Calloni R, Cordero EA, Henriques JA, Bonatto D (2013) Reviewing and updating the major molecular markers for stem cells. *Stem Cells Dev* 22: 1455–1476. doi: [10.1089/scd.2012.0637](#) PMID: [23336433](#)
72. Lecourt S, Marolleau JP, Fromiguet O, Vauchez K, Andriamanalijaona R, Ternaux B, et al. (2010) Characterization of distinct mesenchymal-like cell populations from human skeletal muscle in situ and in vitro. *Exp Cell Res* 316: 2513–2526. doi: [10.1016/j.yexcr.2010.04.020](#) PMID: [20430024](#)
73. McMillan EM, Quadriatero J (2014) Autophagy is required and protects against apoptosis during myoblast differentiation. *Biochem J* 462: 267–277. doi: [10.1042/BJ20140312](#) PMID: [24865278](#)
74. Skop V, Cahova M, Dankova H, Papackova Z, Palenickova E, Svoboda P, et al. (2014) Autophagy inhibition in early but not in later stages prevents 3T3-L1 differentiation: Effect on mitochondrial remodeling. *Differentiation*.
75. Baerga R, Zhang Y, Chen PH, Goldman S, Jin S (2009) Targeted deletion of autophagy-related 5 (atg5) impairs adipogenesis in a cellular model and in mice. *Autophagy* 5: 1118–1130. PMID: [19844159](#)
76. Martinez-Lopez N, Athonvarangkul D, Sahu S, Coletto L, Zong H, Bastie CC, et al. (2013) Autophagy in Myf5+ progenitors regulates energy and glucose homeostasis through control of brown fat and skeletal muscle development. *EMBO Rep* 14: 795–803. doi: [10.1038/embor.2013.111](#) PMID: [23907538](#)

2.2 – Conclusion

La richesse et la fiabilité des résultats produits par ProteINSIDE, à partir d'une liste de 1000 ID choisis aléatoirement pour chaque espèce, est similaire pour les espèces monogastriques, mais meilleures pour les espèces de ruminants à celle des résultats produits par David, BioMyn ou AgBase. Les différences de résultats produits par les 4 outils pour les modules d'analyses communs, la synthèse des informations et la GO, résultent vraisemblablement de différences dans le maintien et la mise à jour des ressources ([Kaspric et al., 2015](#)). En effet, depuis la publication de cette comparaison des résultats, l'outil BioMyn n'est plus maintenu et accessible.

De plus, nous avons vérifié la pertinence biologique des résultats de ProteINSIDE par rapport aux précédentes analyses ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)). Nous avons démontré que ProteINSIDE apporte d'une part des résultats similaires à ceux précédemment publiés et d'autre part de nouvelles connaissances biologiques. Ainsi, dans le TA, ProteINSIDE a identifié de nouveaux termes GO enrichis, tels que des processus liés au métabolisme des lipides (THEM4, ECH1, ACAT2, ECHS1, GPD1 et PCCB, respectivement Acyl-coenzyme A thioesterase, Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, Acetyl-CoA acetyltransferase cytosolic, Enoyl-CoA hydratase, Glycerol-3-phosphate dehydrogenase [NAD(+)] et Propionyl-CoA carboxylase beta chain) et des acides gras (ECHS1, ECH1, PCCB, et THEM4) pour les protéines identifiées à 260 jpc. Ces résultats sont cohérents avec l'augmentation du volume des adipocytes et les activités d'enzymes lipogéniques dans le TA périrénal ([Taga et al., 2011](#); [Taga et al., 2012](#)). Dans le cas du muscle, ProteINSIDE a mis en évidence des termes GO liés à la régulation négative de processus apoptotiques (ALB, ACTC1 et PSMB7, respectivement Albumin, Actin alpha cardiac muscle 1 et Proteasome subunit beta type-7), aux processus de la glycolyse (ENO1, Alpha-enolase) et à l'assemblage des filaments fin d'actine dans les muscles squelettiques (ACTC1) pour les protéines identifiées à 60 jpc. Parmi les résultats originaux apportés par ProteINSIDE, nous proposons 18 protéines potentiellement sécrétées par le TA et 9 sécrétées par le muscle. La présence d'adiponectine (ADIPOQ), une adipokine, dans le muscle fœtal a suscité notre intérêt, car c'est une protéine synthétisée et sécrétée par le TA, donc par les adipocytes intramusculaires fœtaux (qui ne sont pas encore différenciés ; [Bonnet et al., 2010](#)). Elle est connue pour être impliquée dans la différenciation des adipocytes et dans la prolifération, la différenciation et le métabolisme du muscle. La synthèse de cette protéine par le muscle et notamment par le muscle fœtal paraissait surprenante, ainsi une quantification des ARNm correspondant dans le muscle fœtal a confirmé que la protéine n'était pas synthétisée par ce tissu. Elle est toutefois présente dans le muscle fœtal et pourrait impacter sa prolifération et sa différenciation.

Des interactions entre le muscle et le TA sont susceptibles de réguler leurs croissances. Afin d'augmenter nos connaissances sur le dialogue cellulaire et les voies de signalisation impliquées dans la croissance fœtale des deux tissus et leurs interactions, nous avons créé un jeu de données composé des protéines communes aux deux tissus et analysé simultanément les 2 jeux de données. Nous avons identifié 46 protéines présentes à la fois dans le muscle et le TA. L'annotation fonctionnelle des protéines par des termes GO met en évidence des protéines impliquées dans l'apoptose, le cycle cellulaire, la régulation des voies de signalisation impliquant des cytokines ou des interleukines. Nous avons identifié 6 protéines sécrétées qui sont communes aux deux tissus (ADIPOQ, APOA1, AFP, ALB, PDIA3 et SERPINA1, respectivement Apolipoprotein A-I, Alpha-fetoprotein, Protein disulfide-isomerase A3 et Alpha-1-antitrypsin). Avec seulement 46 protéines, le réseau de PPI ne révélait aucun lien d'intérêt. Aussi, nous avons réalisé un réseau étendu de PPI afin

de révéler des interactions entre des protéines communes aux deux tissus et d'autres protéines extérieures au jeu de données et qui pourraient être impliquées dans la régulation de l'ontogenèse des tissus. Parmi ces PPI, nous avons mis en évidence des protéines centrales dans les réseaux qui ne font pas partie des 46 protéines communes et qui sont connues pour être impliquées dans des mécanismes d'autophagie (GABARAPL1, GABARAPL2, GABARAP, MAP1LC3A et MAP1LC3B qui sont des membres de la famille des protéines Atg8, avec respectivement les « Gamma-aminobutyric acid receptor-associated protein-like 1 et 2 » et les « Microtubule-associated proteins 1A/1B light chain 3A et 3B »). L'hypothèse de l'implication du processus d'autophagie dans l'ontogenèse des tissus, est soutenue par des résultats *in vitro* puisque certaines protéines de l'autophagie participent aux processus de différenciation des lignées musculaire C2C12 et adipeuse 3T3-L1 ([McMillan et Quadriatero, 2014](#); [Skop et al., 2014](#)). Il faut cependant préciser que ces protéines extérieures au jeu de données ne sont pas forcément synthétisées par les TA ou les muscles ou pendant la période fœtale, et que l'expression des gènes ou l'abondance de ces protéines durant la période fœtale doit être vérifiée expérimentalement.

Malgré l'apport de connaissances liées à l'application de ProteINSIDE à des données de protéomiques existantes, les 143 protéines adipeuses et les 120 protéines musculaires constituaient une faible part de la caractérisation du protéome fœtal. Elles ont donc été complétées par une nouvelle étude de protéomique.

3 – Caractérisation du protéome fœtal du muscle et du tissu adipeux à 140 jpc

Afin de compléter les connaissances sur le protéome des TA et des muscles, et d'identifier les mécanismes moléculaires qui concourent à la croissance des tissus à la fois par hyperplasie et par hypertrophie, nous avons caractérisé le protéome de deux TA viscéraux et de deux muscles à 140 jpc. En outre et pour identifier des protéines potentiellement clé dans la régulation de la croissance des TA et des muscles, nous avons recherché les protéines dont l'abondance varie en fonction de la nutrition maternelle. Les données concernant l'impact de la nutrition maternelle ont fait l'objet d'une communication au 61^{ème} congrès international sur les sciences et technologies des viandes. La caractérisation des protéomes chez le fœtus fait l'objet d'une publication en préparation.

3.1 – Impact de l'alimentation maternelle sur le protéome des tissus adipeux et musculaires

Introduction

Des données bibliographiques principalement chez le mouton, montrent que l'alimentation des mères au cours de la période fœtale peut avoir des conséquences sur le développement de ces tissus encore visibles au stade adulte ([Bonnet et al., 2010](#)). En effet les rares données disponibles et obtenues essentiellement chez l'ovin, indiquent qu'une sous-nutrition maternelle durant la première moitié de gestation a pour conséquence une augmentation du diamètre des fibres, une diminution du ratio fibres rapides/lentes et une augmentation de l'adiposité des fœtus (pour revue : ([Bonnet et al., 2010](#))). Une augmentation du diamètre des fibres du muscle Infraspinatus a aussi été observée chez des fœtus bovins issus de mères sous-nutries de 30 à 85 jours de gestation. Cependant le diamètre des fibres est diminuée dans le muscle foetal si la sous-nutrition est prolongée jusqu'à 140 jours ou si la sous-nutrition est suivie d'une réalimentation entre 85 et 140 jours ([Gonzalez et al., 2013](#)).

Ces fœtus bovins, différant par des caractéristiques myogéniques, ont été le support d'une étude différentielle des protéomes adipeux et musculaires afin d'identifier des protéines susceptibles de contribuer à une régulation de la myogenèse et de l'adipogenèse à une période clé du développement : avant la deuxième moitié de la gestation. Nous avons expliqué dans l'étude bibliographique que nombre total de fibres musculaires est fixé à environ 180 jpc de la gestation chez le bovin ([Picard et al., 2002](#)). Les événements qui déclenchent l'équilibre entre le nombre de cellules musculaires et adipeuses se produisent pendant la vie fœtale ([Bonnet et al., 2015](#)). Les effets de la manipulation de la nutrition maternelle sur la croissance du TA et du muscle ont été étudiés chez les ovins, mais rarement chez les bovins ([Du et al., 2010](#); [Yan et al., 2013](#)). Les connaissances actuelles montrent que la période qui impacte, par la restriction des nutriments de la mère, sur la croissance musculaire se situe avant la moitié de la grossesse chez les brebis. Il en résulte un diamètre accru des fibres musculaires « rapides », sans changement de diamètre des fibres musculaires « lentes » et une diminution des proportions de fibres musculaires hybrides « rapide et lente » par unité de surface. La dénutrition maternelle, du début de la gestation (28 à 80 jpc), a augmenté la masse grasse fœtale chez les fœtus à court terme ([Bonnet et al., 2010](#)).

L'impact de la restriction alimentaire suivie d'une réalimentation sur le TA et le muscle, ainsi que sur leur équilibre, n'a pas encore été démontré. **L'objectif était d'évaluer les effets d'une restriction suivie ou non par une réalimentation durant la première moitié de gestation, et ceci à la fois sur le muscle et le TA.**

Communication 4

Maternal nutrient restriction followed by realimentation from early to mid-gestation in beef cows impacts proteomes of fetal muscular and adipose tissues

Muriel BONNET^{1, 2}, Nicolas KASPRIC^{1, 2}, Kimberly VONNAHME³, Christiane BARBOIRON^{1, 2}, Didier VIALA^{1, 2, 4}, David CHADEYRON^{1, 2}, and Brigitte PICARD^{1, 2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213, Lempdes, France

3 Department of Animal Sciences, North Dakota State University, Fargo, ND 58108, Etats Unis

4 QuaPA UR 370, INRA, 63122 Saint Genès Champanelle, France

*61th International Congress of Meat Science and Technology (ICoMST), Clermont-Ferrand (France) -
Publié le 23 août 2015 (4 pages).*

MATERNAL NUTRIENT RESTRICTION FOLLOWED BY REALIMENTATION FROM EARLY TO MID-GESTATION IN BEEF COWS IMPACTS PROTEOMES OF FETAL MUSCULAR AND ADIPOSE TISSUES

M. Bonnet^{1,2}, N. Kaspric^{1,2}, K. Vonnahme³, C. Barboiron^{1,2}, D. Viala^{1,2,4}, D. Chadeyron^{1,2},
B. Picard^{1,2}

¹INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

²Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

³Department of Animal Sciences, North Dakota State University, Fargo, ND 58108

⁴QuaPA UR 370, INRA, 63122 Saint Genès Champanelle, France

Abstract – Perturbations of the prenatal environment may influence fetal muscle and adipose tissue (AT) development. This study investigated the effects of nutrient restriction and realimentation from early to mid-gestation on molecular events that underlie muscle and AT growth. On day 30 of pregnancy, 18 multiparous, non-lactating cows were fed at requirement (C; 100% NRC) or restricted (R; 60% NRC). On day 85, cows remained on control (CC; n = 5) or restricted (RR; n = 5) diets, or were realimented to control (RC; n = 5). A proteomic analysis of the *Semitendinosus* and *Longissimus thoracis* muscles as well as perirenal and omental AT from 140 days-old fetuses was conducted using 2-DE and MS. The abundances of 28 and 34 proteins were modified (P<0.10) by maternal nutrition in muscles and AT, respectively. These proteins are involved in glucose or protein metabolisms, in the regulation of cell proliferation or apoptosis, and thus could be related to variations in the number and the size of myofibers and adipocytes. These results are of interest for the fetal programming of myogenesis and adipogenesis in general, and more specifically to the control of muscle and AT growth in meat producing animals.

Key Words – adipose tissue, muscle, growth fetal programming

I. INTRODUCTION

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 is the result of a dynamic balance between the number and size of muscular and adipose cells. The number of muscle fibers is set during gestation [2], implying that any events that trigger a regulation or balance between the number

of muscular and adipose cells must occur and be regulated during fetal life. The effects of manipulating maternal nutrition on the muscle and AT growth have been studied in sheep but rarely in cattle. Current knowledge shows that the most critical period for the impact of maternal nutrient restriction on muscle growth is before the half of pregnancy in ewes. It resulted in an increased diameter of fast myofibers, no change in diameter of slow myofibers and a decreased fast-to-slow myofiber ratio per unit area. Concomitantly maternal undernutrition, during early to mid-gestation (28 to 80 days) increased the fetal fat mass in near-term fetuses (for review see [3]). The molecular mechanisms associated with altered AT growth [4,5] and modified myofibers structures [1,6] begin to be revealed in ruminant fetuses. However the impact of nutrient restriction followed by realimentation on AT and muscle growths, as well as on their balance, remains to be unraveled [2]. Our objectives were to evaluate the effects of nutrient restriction during early gestation followed or not by realimentation to mid-gestation both on AT and muscle growth as well as on the related molecular events.

II. MATERIALS AND METHODS

Perirenal and omental AT, *Semitendinosus* (ST) and *Longissimus thoracis* (LT) muscles were sampled from bovine fetuses at 140 days post-conception (dpc) as previously described [1]. Briefly, Angus cows were randomly assigned to three dietary treatments at 30 days of pregnancy: cows maintained at 100% nitrogen and energy (NE) recommendations for maintenance and fetal growth from 30 to 140 dpc (CC, n= 5), cows

nutrient restricted to 60% NE recommendations until 85 dpc followed by realimentation to 100% NE recommendations up to 140 dpc (RC, n=5) and cows nutrient restricted to 60% NE from 30 to 140 dpc (RR, n= 5).

Proteomic experiments were assayed to identify molecular events modified by maternal nutrition that may affect AT and muscle growth, as well as a balance between their growths. Total protein were extracted as previously described [7], and were separated in a first-dimensional electrophoresis using 5-8 pH range. These proteins were separated in the second dimension on separating 12% polyacrylamide gels and were stained with G250 colloidal coomassie blue. Protein spot detection and volume quantification were realized with SameSpots V4.5 software. We detected 368 muscular and 359 adipose well-resolved protein spots in the gel image. Trypsin-digested proteins were analyzed by online nanoflow liquid chromatography using an Ultimate 3000 RSLC (Dionex, Voisins Le Bretonneux, France) coupled to a LTQ VELOS mass spectrometer (ThermoFisher Scientific, Courtaboeuf, France). MS/MS ions were analyzed using Proteome Discoverer V1.4.1.14 and compared to *Bos taurus* nrNCBI database using MASCOT V 2.3. We have identified 98 adipose and 69 muscular proteins. Protein abundances were analyzed using variance analysis (ANOVA, R software) to compare diets whatever the anatomical site of AT and muscle. Proteins whose abundance changed with diet in CC *versus* RC or RR were considered significant at $P < 0.10$. Data mining of proteomics results was performed using ProteINSIDE (www.proteinside.org) an online tool that we have developed to analyze lists of protein or gene identifiers from well-annotated species (human, rat, and mouse) and ruminants (cow, sheep, and goat) [8]. Briefly, ProteINSIDE gathers biological information stored in well-updated public databases for a systematic and integrative analysis of a protein's biological data, proceeds to annotations according to the Gene Ontology (GO) consortium to identify over expressed functions in a list of identifiers, predicts potentially secreted proteins that mediate signalization between cells or tissues, and searches for proteins interactions to identify proteins contributing to a process or to predict protein function.

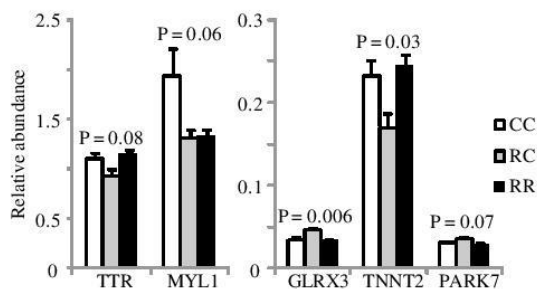
III. RESULTS AND DISCUSSION

At 140 dpc, absolute and relative (to body or eviscerated body weights) weights of perirenal and omental plus mesenteric AT were not modified by maternal nutrition. However an impact of maternal nutrition on muscle growth potential was sustained by a decreased fiber cross-sectional area of the *Infraspinatus* muscle in RC and RR fetuses compared to CC ones [1]. The abundance of 34 adipose (identified in omental or perirenal AT) and 28 muscular (identified in ST or LT muscles) proteins were modified by maternal nutrition (in RC or RR when compared to CC) and 4 (CCT6A, HBBF, PDIA3, RBP4) were retrieved both in AT and muscles. Thanks to the gathering of biological information from the main public databases, ProteINSIDE allows to identify at a glance proteins that could be key in the regulation of tissues growth. As an example for muscular proteins, structural proteins such as TNNT2 and MYL1 were highlighted, as well as proteins such as PARK7 that was proposed to be a regulator of the proliferation of the embryonic muscle cell lineage [9]. Isoforms of troponins, MYL1 and PARK7 were already identified in fetal bovine muscle [9], and the present study revealed that their abundances were regulated by maternal nutrition (Fig. 1).

The muscular proteins were annotated with enriched GO terms related to biological processes such as “muscle filament sliding” which encompassed 3 proteins (ACTC1, MYL1, TNNT2), “tricarboxylic acid cycle” (DLD, PDHB) and “glycolytic process” (ENO1, PGAM1), “negative regulation of apoptotic process” (ANXA5, PARK7, PSMA2), “cellular protein metabolic process” (PDIA3, VBP1, RPS12, CCT6A). In regards to these results, an increase in the abundance of glycolytic proteins was reported between 110 and 180 dpc to sustain myogenesis [9]. Moreover an increase in glycolytic and oxidative metabolisms was repeatedly reported to be associated with muscular hypertrophy in growing cattle [10]. Lastly an inhibition of apoptosis was proposed to regulate the number of muscular cells during fetal bovine myogenesis [9]. Present results show that abundances of proteins involved in these pathways are modified by maternal nutrition, and consequently could induce or be related to variations in the number and the

size of myofibers in bovine fetuses (see the abstract by B. Picard).

Figure 1. Abundance of TTR and GLRX3 in AT and of MYL1, TNNT2 and PARK7 in muscles from fetuses

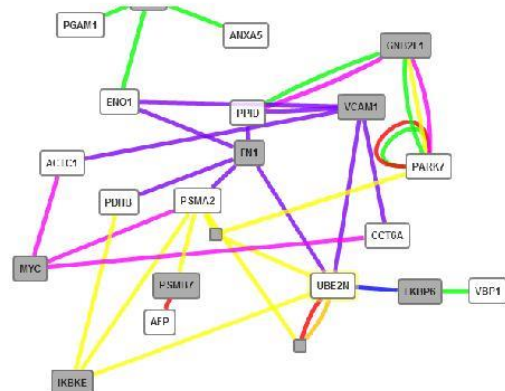


The adipose proteins were annotated with enriched GO terms related to biological processes “protein folding” and “cellular protein metabolic process” (PDIA3, CCT5, CCT3, CCT6A), “arginine catabolic process” (ARG1, DDAH2), “cell proliferation” (FSCN1, SERPINF1), “negative regulation of apoptotic process” (DDAH2, PSMB2). Pathways related to the metabolism of amino acids, as well as to rates of proliferation, apoptosis, and cell survival were already identified during the time course of bovine fetal adipogenesis [7]. However, even if pathways were similar, the majority of the proteins identified in the present study were not identified in our previous one. Among the newly identified proteins, GLRX3 known to regulate the function of the thioredoxin system, was never identified in AT. However, GLRX3 was proposed to protect cells against oxidative stress and to enhance cell cycle progression during late mitosis in mice embryos [11]. The thyroid hormone-binding protein, transthyretin (TTR) involved in the transport and availability of thyroid hormones to peripheral tissues was previously identified in fetal bovine AT [7] (probably as a blood protein) and was shown to have a modified abundance by maternal nutrition in the present study (Fig. 1). The thyroid hormones, thyroxine (T4) and triiodothyronine (T3), are essential for normal growth and development of the fetus [12,13]. Availability of T4 has been shown to regulate glucose oxidation in AT of fetal pigs. Plasma T3 concentrations were positively related to a reduced efficiency of ATP production in AT of fetal sheep [12]. The variation in the abundance of TTR by maternal nutrition

may modify the availability in thyroid hormones for fetal AT and consequently the metabolisms that sustain fetal adipogenesis.

To complement the results provided by the GO annotation, we looked for additional pathways and central proteins that could regulate myogenesis and adipogenesis, by analyzing the protein-protein interactions (PPI). We provide here results obtained for muscles (Fig.2).

Figure 2. Protein-protein interaction network that includes proteins identified in bovine fetal muscles (white boxes) and proteins external to our dataset (grey boxes). The network was filtered using high values of betweenness (1500, which quantifies how frequently a node is on the shortest path between every pair of nodes for detecting bottlenecks in a network) and closeness (0.1, which quantifies how short are the minimal paths from a given node to all others: a high closeness indicates that a node is close to the topological center of the network) centralities. Edges are experimental methods used to identify protein-protein interactions (one color by methods)



By querying databases IntAct, UniProt, and BioGrid (chosen for their high data curation) we identified 19 muscular proteins that were experimentally proven in Human to interact with 349 proteins that were outside of our dataset. The central (proteins that link the highest number of proteins or proteins that are involved in the shortness paths) proteins of this network were sorted (Fig. 2). The role of PARK7 for myogenesis regulation is sustained by the present analysis, and could be related to the positive effect of Park7 on myofiber diameter and sarcomeric myosin expression reported *in vitro* [14] and the lower proportion of small fiber in RC group (see abstract by B. Picard). The role of enzymes related to

glycolysis (ENO1, PGAM1) and tricarboxylic acid cycle (LDHB) was also revealed by PPI network analysis. As new putative central proteins for myogenesis that were identified as nutritionally regulated in fetal muscles, we highlighted CCT6A, VBP1, UBE2N, that were never reported as muscular proteins and their role in myogenesis remains to be studied.

Lastly, PDIA3 in muscle, as well as PDIA3, TTR and SERPINF1 in AT, were identified as secreted proteins. Whether these proteins contribute to the regulation of myogenesis and adipogenesis or to the balance between these processes remains to be unraveled.

IV. CONCLUSION

This *in vivo* investigation of nutritionally regulated proteome during bovine myogenesis and adipogenesis provides new perspectives on the regulation of muscle and AT development. We have identified proteins with an abundance modified in response to maternal nutrition that could be key in the regulation of tissues growth, which remains to be proven by additional experiments.

ACKNOWLEDGEMENTS

The authors thank the regional council of Auvergne in France and APIS-GENE for funding the N. Kaspric's PhD grant as well as the INRA PHASE department of research for funding the assays.

REFERENCES

- Gonzalez, J. M., Camacho L. E., Ebarb S. M., Swanson K. C., Vonnahme K. A., Stelzleni A. M., Johnson S. E. (2013). Realimentation of nutrient restricted pregnant beef cows supports compensatory fetal muscle growth. *Journal of Animal Science* 91:4797-4806.
- Picard, B., Lefaucheur L., Berri C., Duclos M. J. (2002). Muscle fibre ontogenesis in farm animal species. *Reproduction, Nutrition, Development* 42:415-431.
- 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* 4:1093-1109.
- Bispham, J., Gardner D. S., Gnanalingham M. G., Stephenson T., Symonds M. E., Budge H. (2005). Maternal nutritional programming of fetal adipose tissue development: Differential effects on messenger ribonucleic acid abundance for uncoupling proteins and peroxisome proliferator-activated and prolactin receptors. *Endocrinology* 146:3943-3949.
- Bispham, J., Gopalakrishnan G. S., Dandrea J., Wilson V., Budge H., Keisler D. H., Broughton Pipkin F., Stephenson T., Symonds M. E. (2003). Maternal endocrine adaptation throughout pregnancy to nutritional manipulation: Consequences for maternal plasma leptin and cortisol and the programming of fetal adipose tissue development. *Endocrinology* 144:3575-3585.
- Quigley, S. P., Greenwood P. L., Kleemann D. O., Owens J. A., Bawden C. S., Nattrass G. S. (2015). Myogenesis in small and large ovine fetuses at three stages of pregnancy. *Animal Production Science* 55:207-212.
- Taga, H., Chilliard Y., Meunier B., Chambon C., Picard B., Zingaretti M. C., Cinti S., Bonnet M. (2012). Cellular and molecular large-scale features of fetal adipose tissue: Is bovine perirenal adipose tissue brown? *Journal of Cellular Physiology* 227 1688-1700
- Kaspric, N., Picard B., Reichstadt M., Tournayre J., Bonnet M. Proteinside for easy investigation of proteomics data from ruminants: Application to mine the proteome of adipose and muscle tissues in bovine fetuses. *PLoS ONE*:submitted.
- Chaze, T., Meunier B., Chambon C., Jurie C., Picard B. (2008). In vivo proteome dynamics during early bovine myogenesis. *Proteomics* 8:4236-4248.
- Picard, B., Cassar-Malek I., Guillemin N., Bonnet M. Quest for novel muscle pathway biomarkers by proteomics in beef production. In: *Comprehensive biotechnology* (second edition), Moo-Young M (Ed.), Academic Press, 2011, pp 395-405.
- Cheng, N. H., Zhang W., Chen W. Q., Jin J., Cui X., Butte N. F., Chan L., Hirschi K. D. (2011). A mammalian monothiol glutaredoxin, grx3, is critical for cell cycle progression during embryogenesis. *FEBS Journal* 278:2525-2539.
- Forhead, A. J., Fowden A. L. (2014). Thyroid hormones in fetal growth and parturition maturation. *Journal of Endocrinology* 221:R87-R103.
- Cassar-Malek, I., Picard B., Kahl S., Hocquette J. F. (2007). Relationships between thyroid status, tissue oxidative metabolism, and muscle differentiation in bovine fetuses. *Domestic Animal Endocrinology* 33:91-106.
- Yu, H., Waddell J. N., Kuang S., Bidwell C. A. (2014). Park7 expression influences myotube size and myosin expression in muscle. *PLoS ONE* 9:e92030.

Conclusion

L'abondance de 34 protéines adipeuses (identifiées dans TA omental ou péri-rénale) et 28 musculaires (identifiées dans les muscles ST ou LT) ont été modifiées par la nutrition maternelle (régimes réduit-complet ou réduit-réduit par rapport à un régime complet total ; cf. Matériels et Méthodes 1.1, [Figure 2](#)). De plus, 4 protéines ont été identifiées à la fois dans les deux TA et les deux tissus musculaires (CCT6A, HBBF, PDIA3 et RBP4). Les protéines musculaires ont été annotées avec des termes de GO enrichis liés aux processus biologiques tels que « muscle filament sliding » qui englobait 3 protéines (ACTC1, MYL1 et TNNT2), « tricarboxylic acid cycle » (DLD et PDHB), « glycolytic process » (ENO1 et PGAM1), « negative regulation of apoptotic process » (ANXA5, PARK7 et PSMA2) et « cellular protein metabolic process » (PDIA3, VBP1 RPS12 et CCT6A). Les protéines adipeuses ont été annotées avec des termes de GO enrichis liés aux processus biologiques « protein folding » et « cellular protein metabolic process » (PDIA3, CCT5, CCT3 et CCT6A), « arginine catabolic process » (ARG1 et DDAH2), « cell proliferation » (FSCN1 et SERPINF1) et « negative regulation of apoptotic process » (DDAH2 et PSMB2). Au moins dans le muscle *Semitendinosus*, certaines des protéines identifiées à 140 jours et dont l'abondance varie pourraient avoir contribué au retard de la différenciation des fibres IIA ou à la prolifération inférieure de la troisième génération de myoblastes ([Picard et al., 2015](#)) chez les fœtus de 254 jours issus de vaches restreintes.

3.2 – Description des protéomes adipeux et musculaire à 140 jpc

Introduction

Les études protéomiques visant à caractériser l'ontogenèse du muscle ([Chaze et al., 2008](#); [Chaze et al., 2009](#)) et du TA ([Taga et al., 2012](#)) ont permis de constituer deux protéomes composés de peu de protéines : 120 à 143 protéines, respectivement pour le muscle et le TA. En effet, dans ces études les analyses de spectrométrie de masse par MALDI-ToF n'ont sélectionné qu'une seule protéine par spot, écartant ainsi un grand nombre de protéines des analyses. A l'heure actuelle les analyses de spectrométrie de masse par LC-MS/MS augmentent la sensibilité de la détection et identifient de nombreuses protéines par un spot. Pour assurer la fiabilité de l'identification nous avons sélectionné les protéines identifiées par au moins 3 peptides.

L'objectif était de décrire le plus largement possible les protéomes de chaque tissu à 140 jpc afin d'identifier les processus cellulaires qui accompagnent la croissance des tissus et plus particulièrement leur hyperplasie, et de rechercher des protéines impliquées dans le dialogue entre le TA et le muscle. Pour atteindre cet objectif, nous avons utilisé l'outil ProteINSIDE, pour la fouille de données et l'identification des protéines sécrétées. Pour détecter les protéines impliquées dans le dialogue, un réseau d'interactions moléculaires entre les protéines de surface cellulaire d'un tissu et les protéines potentiellement sécrétées par l'autre tissu a été créé. Pour cette analyse, les protéines de surface référencées dans la base de données Cell Surface Protein Atlas (les protéines proviennent de l'espèce humaine ; [Bausch-Fluck et al., 2015](#)) ont été utilisés. Nous avons comparé nos jeux de données avec ces protéines pour lister 106 et 55 protéines de surface respectivement pour le TA et le muscle. En complément, nous avons identifié les protéines prédites comme sécrétées par ProteINSIDE. Il s'agit de deux listes de 67 et 34 protéines respectivement pour le TA et le muscle.

ARTICLE 5

How mapping the muscle and adipose tissues proteomes could help in the understanding of the “adipose-muscular” cross-talk during fetal growth?

Nicolas KASPRIC^{1, 2}, Brigitte PICARD^{1, 2}, Kimberly VONNAHME³, Christophe CHAMBON^{1, 2} and Muriel BONNET^{1, 2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213, Lempdes, France

3 Department of Animal Sciences, North Dakota State University, Fargo, ND 58108, Etats Unis

En préparation pour une soumission dans le journal Molecular and Cellular Proteomics.

Title (2 lignes max)

How mapping the muscle and adipose tissues proteomes could help in the understanding of the “adipose-muscular” cross-talk during fetal growth?

Nicolas Kaspric^{1,2}, Brigitte Picard^{1,2}, K. Vonnahme³, Christophe Chambon⁴ and Muriel Bonnet^{1,2}

1 INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

2 Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

3 Department of Animal Sciences, North Dakota State University, Fargo, ND 58108

4 QuaPA UR 370, INRA, 63122 Saint Genès Champanelle, France

Running title (60 caractères max)

Mapping the fetal muscle and adipose tissues proteomes.

INTRODUCTION

An adequate development of adipose tissue relatively to muscle (i.e. lean-to-fat ratio) is central both for issues related to the human health and to the meat industry. Indeed an excessive fat accumulation is related to obesity and related pathologies such as metabolic syndrome, insulin resistance or cardiovascular diseases. Moreover, an excessive fat accumulation within the body or muscles of species of agronomic interest is among the major determinant of economic values in the meat industry (Bonnet *et al.*, 2010; Hausman *et al.*, 2014). The lean-to-fat ratio is the result of a dynamic balance between the number and volume of muscular and adipose cells respectively. The fetal life is the unique period during which the muscular masses grow both by an increase in the number of cells (hyperplasia) and by cell enlargement (hypertrophy) in large mammals such as Human and bovine. Indeed, the total number of muscle fibers is set by the end of the second trimester of gestation in bovine (Picard *et al.*, 2002) and in Human (Draeger *et al.*, 1987). It results

from the proliferation and differentiation of the successive generations of myoblasts. Conversely, the number of adipocytes is set by birth or by early adulthood, depending on the anatomical location of the adipose tissue both in human (Spalding *et al.*, 2008) and bovine (Bonnet *et al.*, 2010). Thus, the fetal life is a key period to understand the mechanisms underlying the dynamic balance between the number and volume of muscular and adipose cells in large mammals. However, this implies to know the cellular and molecular events underpinning the fetal growth of adipose tissue and muscle. In muscles, the molecular pattern contributing to the proliferation and differentiation of the different successive generations of myoblasts has been studied using genomic tools (Chaze *et al.*, 2008; Chaze *et al.*, 2009; Lehnert *et al.*, 2007; Sudre *et al.*, 2003). This has revealed that the total number of fibers is controlled by a balance between proliferation and apoptosis, and that complex modifications in proportions of protein isoforms govern contractile and metabolic differentiation during the last third of gestation. In adipose tissue, the total number of perirenal adipocytes increases strongly until the end of the fetal life in bovine (Landis *et al.*, 2002; Taga *et al.*, 2012). The volume of adipocytes increases mainly during the last 40 days of pregnancy in bovine from lean breeds (Taga *et al.*, 2012) while this occurs earlier (-110 to -40 days before birth) in bovine from fat breeds. However, the molecular patterns underlying these cellular events of tissues growth are scarce. A unique proteomic study has identified 143 proteins in the bovine fetal AT, and some of these proteins were proposed to regulate adipocyte precursor proliferation by controlling cell cycle progression and/or apoptosis or delaying PPAR gamma-induced differentiation (Taga *et al.*, 2012), a master pathway of adipogenesis. A first attempt to understand the growth of muscle relatively to adipose tissue using available data (Kaspric *et al.*, 2015) was partly unsuccessful because of an incomplete description of fetal tissues proteomes. An in depth exploration of both adipose and muscular proteomes during fetal life is thus of critical importance for an improved understanding of the mechanisms underlying the dynamic balance between muscle and AT growth. Such balance may occur through tissue-specific signaling pathways and through the secretion of proteins by adipose tissues (adipokines) and skeletal muscle (myokines), putatively regulating their growth in a reciprocal manner. The mechanisms underlying

the “adipose-muscular” cross-talk are however poorly known (Argiles *et al.*, 2012; Bonnet *et al.*, 2010).

The main objective of the present study was to characterize fetal adipose and muscular proteomes and to identify proteins that may contribute to the regulation of the balance between muscle and AT growth and/or sustain an “adipose-muscular” cross-talk. In the present study we explored, for the first time on a large scale, the proteome of both adipose and muscular tissues in fetuses. The fetal age of 140 days post conception (dpc) was chosen because both adipose and muscular tissues grow by hyperplasia and hypertrophy (Bonnet *et al.*, 2010). Lastly, proteomics data have been mined with the web service ProteINSIDE, dedicated to the analysis of ruminants omics data (Kaspric *et al.*, 2015). Moreover, the proteomics data were analysed relatively to the known cell surface subproteome, or surfaceome (Bausch-Fluck *et al.*, 2015) in order to identify proteins that could be involved in the “adipose-muscular” cross-talk. The original insights provided both by the proteomics data and the data mining methods constitute a starting point in our understanding of the growth of muscle relatively to the adipose tissue.

EXPERIMENTAL PROCEDURES

1. Muscles and adipose tissues sampling

Fifteen fetuses were generated by artificial insemination of multiparous crossbred beef cows (predominately of Angus breeding) at the NDSU Beef Research and Teaching Unit (Fargo, ND) as previously described (Gonzalez *et al.*, 2013). Fetuses were collected at 140 days post-conception (dpc) and perirenal and omental adipose tissues, *Semitendinosus* (ST) and *Longissimus thoracis* (LT) muscles were sampled. Tissues were stored at -80°C pending protein extraction.

II. Large-scale proteomic analysis

Total proteins were extracted as previously described from AT (Taga *et al.*, 2012) and from muscles (Bouley *et al.*, 2004). Proteins were separated by 2-DE electrophoresis performed according to Bouley *et al.* (Bouley *et al.*, 2004). Briefly, for the IEF, proteins were solubilized in a rehydration solution (8M urea, 1 M thiourea, 0.28% DTT, 2% CHAPS, 2% IPG buffer pH 4–7 and 0.01% CBB G-250). IPG strips covering a pH range of 5–8 were loaded with 700 µg of proteins and were subjected to IEF (73.5 kVh) in a Multiphor II (Amersham Biosciences, France) gel apparatus at 20.5°C. After completion of the IEF, proteins on the strip were equilibrated for 15 min in a buffer containing 6M urea, 1% DTT, 30% glycerol, 50 mM Tris base, 2% SDS and DTT, and then for an additional 15 min in the same solution except that DTT was replaced by 5% iodoacetamide. 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, USA). Gels were stained with G250 colloidal coomassie blue. Protein spot detection was realized with SameSpots V4.5 software (Nonlinear Dynamics, Newcastle, UK). Protein spots from the two adipose tissues (30 gels) or the 2 muscles (30 gels) were matched to a reference gel chosen between the 30 gels from one tissue type. We detected 368 muscular and 359 adipose well-resolved protein spots. A maximum of protein spots were sampled for identification.

The peptides mixtures were analyzed by nano-LC-MS/MS using an Ultimate 3000 system (Dionex, Voisins le Bretonneux, France) coupled to an LTQ velos (Thermo scientific) with a nanospray ion source (Bruker Daltonics). Eight microliter of each peptide samples was loaded on a C18 pre-column (300-µm inner diameter × 5 mm; Dionex) at 30 µl/min in 97.9% H₂O, 2% acetonitrile, and 0.1% trifluoroacetic acid. After 6 min of desalting, the pre-column was switched on line with the analytical C18 column (75 µm inner diameter × 15 cm, 3µm; Acclaim C18 Pepmap RSLC) equilibrated in 96% solvent A (99.5% H₂O, 0.5% formic acid) and 4% solvent B (80% acetonitrile, 19.5% H₂O, 0.5% formic acid). Peptides were eluted using a 4 to 90% gradient of solvent B during 29 min at a 300

nl/min flow rate. The eluate was electrosprayed into the mass spectrometer through the nanospray ion source. The LTQ velos was operated in a CID top 10 data dependent mode. Dynamic exclusion was enabled (2 repeat counts, 30 s repeat duration and 60 s exclusion duration).

At the end of each LC-MS/MS analysis, the raw files were loaded into the Protein Scape software (v 3.1, Bruker, Breme, de) and identification was performed with Mascot (<http://www.matrixscience.com>). The database used for the search was UniProt Bos taurus (24,843 sequences), with the following parameters: peptide mass tolerance was set to 500 ppm, fragment mass tolerance was set to 0.2 Da, 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 > 0.05 with a False Discovery Rate (FDR) < 5%).

III. Bioinformatic analysis using ProteINSIDE

Proteomics data were mined with ProteINSIDE (Kaspric et al., 2015). The uniprot identifiers of the bovine proteins identified in adipose and muscular were automatically translated into human identifiers using ortologous and homologous links known between the two species and proposed by ProteINSIDE. Identifiers for adipose and muscular proteins were launched independently for a basic analysis (default options). The basic analysis aims to gather available biological information stored in public databases (NCBI (Coordinators, 2015) and UniProt (UniProt, 2015)) for a protein, and to provide annotations according to the Gene Ontology (GO; (Gene Ontology, 2015)). It also predicts proteins that are secreted to search for proteins that mediate signalization between cells or tissues. For this prediction ProteINSIDE used the algorithm SignalP (Petersen *et al.*, 2011), and the predictions were verified both by GO terms annotations related to extracellular component and the prediction of cellular location thank to TargetP (Emanuelsson *et al.*, 2000). Protein-protein interactions (PPI)

between the proteins of a data set were identified thanks to the PPI referenced UniProt (UniProt, 2015), IntAct (Orchard et al., 2014), and BioGrid (Chatr-Aryamontri et al., 2015) databases. As already described (Kaspric et al., 2015), these PPI databases were chosen as a default option in the PPI module of analysis of ProteINSIDE because they are daily updated and reviewed by curators. Moreover they are included in curation processes of the IMEx (that ensures reliable interactions data using experts and curation rules shared between many interaction databases (Orchard et al., 2012)) or MIMIx (a guideline of the minimum information required for reporting a molecular interaction experiment, thus advising the user on how to use the interaction data (Orchard et al., 2007)) project. In order to extend the adipose and muscular proteomes, PPI were searched between proteins of our experimental datasets and proteins from the human proteome in the same three PPI databases, using the custom analysis of proteINSIDE.

IV. Data mining to identify links between cell surface proteins and secreted proteins

Data from the fetal proteomes characterised in the present study were merged with previous data from the fetal muscular (Chaze *et al.*, 2008; Chaze *et al.*, 2009) and adipose (Taga *et al.*, 2012) proteomes. To date available data have allowed to list 801 and 556 proteins present in the bovine fetal adipose or muscular tissues, respectively. Among these fetal proteins, we searched for first cell surface proteins and second putatively secreted proteins. To identify cell surface protein, we have merged lists of known cell surface proteins using data from the Cell Surface Protein Atlas Database (Bausch-Fluck et al., 2015), cell surface subproteome data from embryonic mouse stem cell (Gu et al., 2010) and mass spectrometry-based cell surface capturing data from the muscular cell line C2C12 (Grundy et al., 2009). A list of 1492 cell surface proteins was established and was used for a comparison with the lists of adipose and muscular fetal proteomes. To search for proteins secreted either by fetal adipose tissue or muscle, the 801 and 556 proteins were launched in ProteINSIDE in order to predict secreted proteins as previously described (Kaspric *et al.*, 2015). Interactions between

secreted proteins of a tissue and cell surface proteins from the other tissue, were searched with the PPI module of ProteINSIDE. Lastly, we have created an overview of protein interactions between tissues using Pathway Studio software (Nikitin et al., 2003). We inputted in Pathway Studio datasets made of secreted proteins of a tissue and cell surface proteins of the other tissue. We have created networks of molecular interactions that highlighted regulations of gene expression that were referenced for adipose or muscular cells. Lastly and to ascertain the biological significance of these interactions in the context of adipose tissue and muscle biology, we have verified the synthesis of a protein by a tissue by using the experimental results reported in Expression Atlas database (Petryszak et al., 2014).

RESULTS

I. Fetal bovine muscle and adipose tissues proteomes

A total of 752 and 514 bovine proteins were identified in AT and muscle, among them 338 proteins were found in the two tissues. As genes related by evolution, particularly orthologs, usually have conserved function between vertebrates, and because human is the most annotated species, human orthologs of bovine proteins were considered in order to further investigate their biological functions. Thus, 705 and 490 bovine proteins from AT and muscles respectively were assigned a human orthologs, thank to HomoloGene database used by ProteINSIDE, and can be viewed in the supplemental Table S1. Among them 331 proteins were found in the two tissues. The human orthologs were subjected to the bioinformatic data mining.

I.1. Functionnal Enrichment for Protein Sets

Over the proteins identified, 338 bovine proteins or 331 humans ortologs were found both in adipose tissues and muscles. Consequently, similar enriched (p-value < 0.001) GO terms related to biological processes were found both for muscle (Figure 1) and for AT (Figure 2). The shared GO terms have been grouped in four main functional categories: cell proliferation, cell differentiation, Wnt signalling and Apoptosis.

GO terms related to cell proliferation were reported in Figure 1 for muscle and in Figure 2 for AT. They annotated 31 muscular and 43 adipose proteins, among them 21 proteins were identified in both tissues (COPS8, CRKL, CTBP1, DNAJA2, FSCN1, JUP, KRT16, KRT6A, MAP2K1, NCK2, NES, NUDT16, OGN, PA2G4, PDAP1, PDXK, RPRD1B, SERPINF1, TPD52L2, TXN and WARS). Within muscle, the identified proteins that could be involved in the overall cell proliferation regulation are JUP (Junction plakoglobin), TPD52L2 (Tumor protein D54), and RBFOX2 (RNA binding protein fox-1 homolog 2), while others proteins were proposed to be involved either in positive (NUDT16, RPRD1B, KRT6A, DNAJA2, MVD, EIF5A, and CRKL) or negative (MAP2K1, COPS8, FABP3, CTBP1, NCK2, and WARS) regulation of the cell proliferation. Among the 31 muscular proteins that were proposed to be involved in cell proliferation, we have also identified RBP4 (Retinol-binding protein 4) and OGN (Mimecan) that were described in contexts spanning negative regulation of smooth and cardiac muscle proliferation, as well as the protein NES (Nestin) related to the proliferation of stem cells. Within AT, the identified proteins that could be involved in the overall regulation of cell proliferation are TPD52L2 and JUP, and additional proteins were involved in the positive (NPM1, NUDT16, PRDX3, RPRD1B, CRKL, KRT6A, DNAJA2, CNBP, and CAPNS1) or its negative (SOD2, CTBP1, COPS8, WARS, NCK2, NPM1, CSNK2B, MAP2K1, PHB, CTBP2, ALDH1A2, and PLG) regulation. Among the 43 adipose proteins that were proposed to be involved in cell proliferation, we have also identified PBLD (Phenazine biosynthesis-like domain-containing protein 1), CDKN1C (Cyclin-dependent kinase inhibitor 1C), TGFBR3 (Transforming growth factor beta receptor type 3), and KRT4 (Keratin, type II cytoskeletal 4) that have been described in the context of negative regulation of epithelial cell proliferation, SERPINF2 (Alpha-2-antiplasmin) and ADIPOQ (Adiponectin) or OGN that were reported

in a context of positive and negative regulation of muscle cell proliferation, respectively; and NES involved in the proliferation of stem cell.

GO terms related to cell differentiation have annotated 11 muscular (Figure 1) and 25 adipose (Figure 2) proteins. Within muscle, PPP2CA (Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform) was annotated with the GO term cell differentiation and PA2G4 (Proliferation-associated protein 2G4) or ANP32B (Acidic leucine-rich nuclear phosphoprotein 32 family member B) with GO terms relative to either positive or negative regulation of cell differentiation. We have also identified SYNE1 (Nesprin-1), MAP2K6 (Dual specificity mitogen-activated protein kinase kinase 6), and PTBP1 (Polypyrimidine tract-binding protein 1) that have been described in the context of muscle differentiation, ANKRD2 (Ankyrin repeat domain-containing protein 2) and HIF1AN (Hypoxia-inducible factor 1-alpha inhibitor) that were reported to regulate the differentiation of myoblasts, and C1QTNF3 (Complement C1q tumor necrosis factor-related protein 3) and CTBP1 (C-terminal-binding protein 1) that were proposed to regulate the differentiation of adipocytes. Within adipose tissue, CLIC4, KRT6A, ADIPOQ, CTBP2, CTBP1 and JUP were annotated with GO terms related to brown or white adipocyte differentiation, or the acquisition of specific characteristics of adipocytes. Among the adipose proteins, we have also identified CDC42 (Cell division control protein 42 homolog), RBM4 (RNA-binding protein 4), and MAPK14 (Mitogen-activated protein kinase 14) that were described in a context related to the positive regulation of muscle cell differentiation.

GO terms related to Wnt Signaling pathway have annotated 16 muscular (Figure 1) and 21 adipose (Figure 2) proteins. The muscular proteins (PSMA1, PSMA6, PSMA7, PSMB2, PSMB3, PSMB4, PSMB6, PSMC1, PSMC6, PSMD7, PSMD9, PSME1, PSME2, UBA52) and the adipose protein (PSMA1, PSMA2, PSMA6, PSMB1, PSMB2, PSMB3, PSMB4, PSMB6, PSMB7, PSMC1, PSMC2) were annotated with the GO terms relative both to the positive and negative regulation of the Wnt signaling pathway. Most of these proteins were involved in the proteasome complex. An additional protein

JUP (Junction plakoglobin) identified in the AT and muscle was annotated specifically with the GO term relative to the positive regulation of the Wnt signaling pathway. PPP2CA identified in AT (Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform) and TBL1XR1 identified both in AT and muscles (F-box-like/WD repeat-containing protein TBL1XR1) were reported to be involved in the Wnt signaling pathway and its regulation.

The apoptosis was the most annotated biological process both in the fetal muscular (56 proteins, figure 1) and adipose (71 proteins, figure 2) proteomes. Both in muscle and adipose tissue most of the proteins are involved in the negative regulation of apoptotic process, since 34 muscular (THBS1 PSMA6 PARK7 PSME2 ANXA5 PA2G4 GOLPH3 PSMD9 UBA52 PSMB4 ANXA1 PSMD7 PSME1 PSMB6 GLO1 PSMA7 HNRNPK HSPD1 HSPA5 BAG3 DDAH2 PAK2 TPT1 LTF PSMC6 ANXA4 PSMC1 PSMB2 ARHGDI A HSPB1 PSMB3 PRDX2 PSMA1 ALB) and 43 adipose (PRDX3 HSPB1 PSMD8 PSMB3 ANXA1 FABP1 ANXA5 DDAH2 PSMC2 PSMD5 PSMA2 PSMB2 TPT1 THBS1 KRT18 PSMD13 PSME2 HNRNPK PRDX2 PSMA1 ARHGDI A GLO1 PSMB4 PSMA6 HSPA5 DDX3X PSME1 NPM1 YWHAZ ALB FHL2 CAT PSMC1 FLNA PSMB6 GSTP1 PSMB7 CFL1 PARK7 PSMB1 PSMD10 PSMD9 PA2G4) proteins were annotated with this GO term. Only 10 muscular (HSPD1 PPID UBA52 KNG1 BIN1 FADD HMGB1 SEPT4 SOD1 HTRA2) and 7 (DDX3X KNG1 PPID SOD1 BID AIFM1 HMGB1) adipose proteins were annotated by the GO term positive regulation of apoptotic process. NPM1 (Nucleophosmin) and APOH (Beta-2-glycoprotein 1) were identified both in muscle and AT and were previously reported in the context of the negative regulation of apoptotic process, and for APOH particularly in smooth muscle cell apoptotic process (Figure 1).

GO terms relative to the functional category “muscle development” were enriched in the muscle dataset, and they annotated 31 proteins (Figure 1). These proteins have been described in contexts spanning the organization of the muscle (FHL1, SRI, LMNA, MYH3, FHL3, ITGB1BP2, LAMA2 and ANKRD2), the assembly of actin and myosin filaments (MYL1, ACTC1, ACTN2, TNNT3, MYL6B, TTN, TPM2, TCAP, MYL4, TNNT2, ACTA1, MYH3, MYH8, MYH2, TNNT1, MYL6), and in

the acquisition of muscle contraction function (MYL6B, DES, MYH2, ACTG2, MYL1, GTAG , TTN, MYH13, MYLPF, MYH8, MYL6, TPM2, ACTA1, ACTA2, TMOD2 and ANKRD2). Other proteins were involved in specific muscle contraction function, such as the acquisition of the contraction of skeletal muscle (STAC3, TNNT3, MYH3, TCAP, TNNT1, TNNT1, and MYH8), striated muscle (CASQ2, TTN, and ALDOA), and smooth muscle (KNG1).

GO terms relative to the functional category “lipid metabolism” were enriched in the AT dataset, and they annotated 10 proteins (Figure 2) mainly involved in the fatty acid oxidation (ECHS1, ACADVL, ACADL, HADHB, PCCB, HADHA, and ADIPOQ) and the negative regulation of this pathway (ERLIN2, ACADVL, and ACADL).

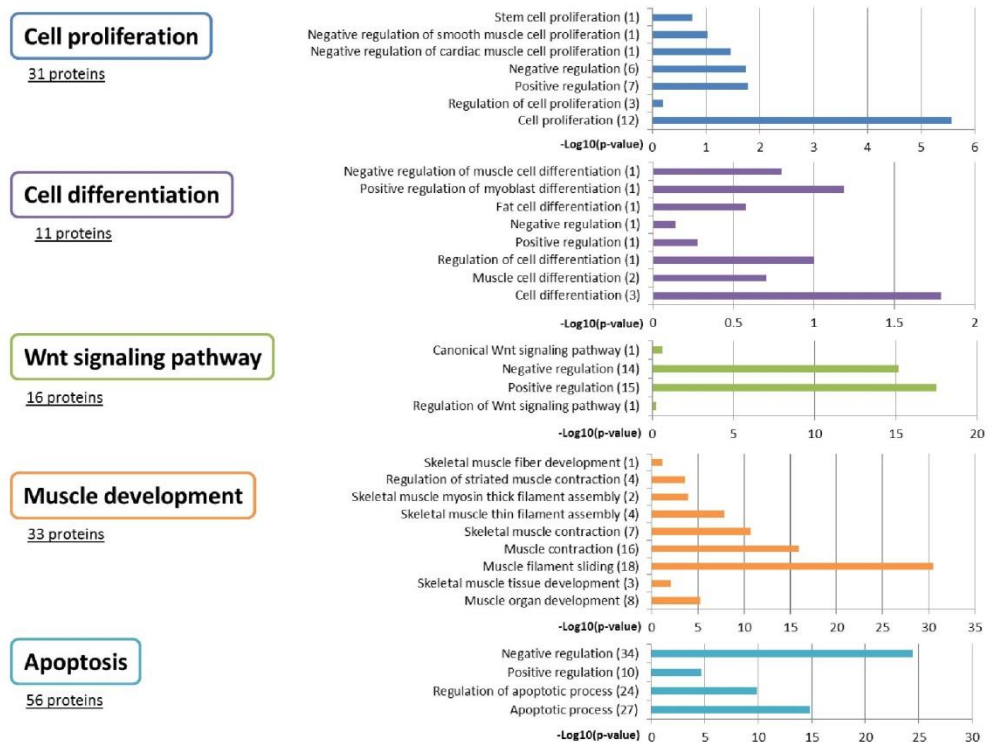


Figure 1: Focus on the most significantly enriched GO terms that annotated proteins identified in the fetal muscle at 140 dpc age. GO terms has been grouped within major biological functions. Graphics

are made using the minus decimal logarithm of the p-value provided by ProteINSIDE. Numbers in brackets after the GO terms are the number of annotated proteins by the GO term.

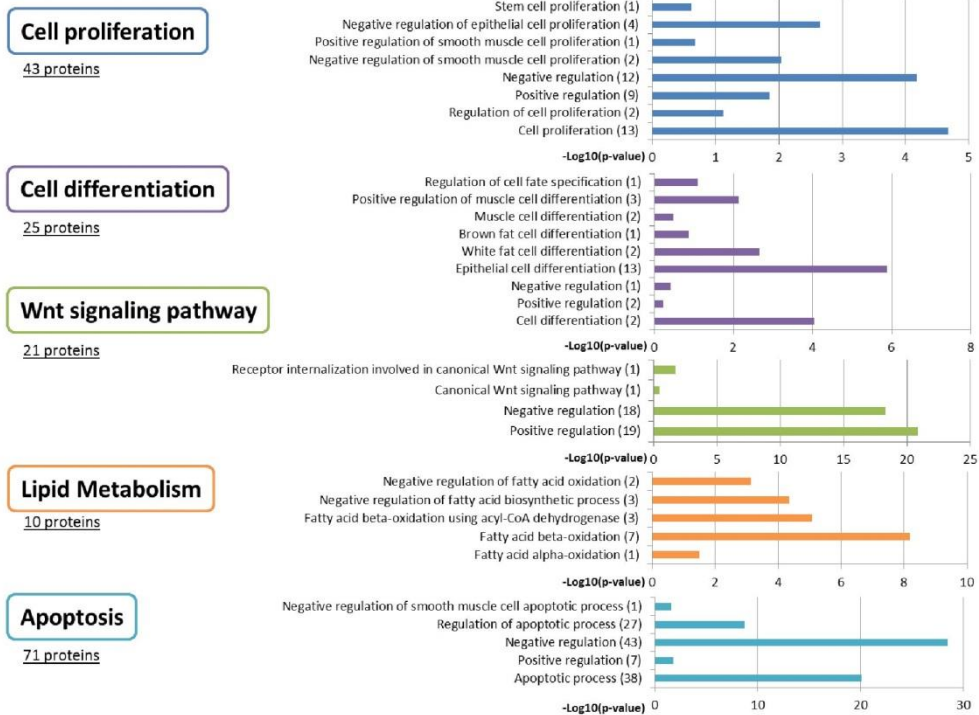


Figure 2: Focus on the most significantly enriched GO terms that annotated proteins identified in fetal AT at 140 dpc age. GO terms has been grouped according to major biological functions. Graphics are made using the minus decimal logarithm of the p-value provided by ProteINSIDE annotation. Numbers in brackets after the GO terms are the number of annotated proteins by the GO term.

1.2. Prediction of secreted proteins

ProteINSIDE has predicted 47 and 93 proteins as secreted thanks to a signal peptide by the fetal muscle and AT, respectively (Table 1). These predictions were confirmed for 28 muscular and 54 adipose proteins according to the subcellular location declared in Uniprot, GO terms annotations

relative to cellular component, and the subcellular location predicted by TargetP provided by ProteINSIDE . We have to note that 36 were predicted as secreted both by AT and muscles and that 24 predictions among the 36 were validated by the three methods proposed by ProteINSIDE (Table 1). A GO annotation of the proteins potentially secreted by the fetal muscle has highlighted proteins which have been described in contexts spanning muscle proliferation (RBP4 and OGN, Retinol-binding protein 4 and Mimecan, respectively), muscle development (RBP4, GSN, and LAMA2, Gelsolin and Laminin subunit alpha-2, respectively), fat cell differentiation and the regulation of adiponectin secretion (C1QTNF3, Complement C1q tumor necrosis factor-related protein 3). The GO annotations of the proteins potentially secreted by the fetal AT has highlighted proteins that could be involved in AT differentiation and metabolism (ADIPOQ, Adiponectin), in muscle proliferation (OGN, ADIPOQ, RBP4, TGFBR3, Transforming growth factor beta receptor type 3, and SERPINF2 Alpha-2-antiplasmin) or development (GSN, RBP4, and COL6A3, Collagen alpha-3(VI) chain).

Because some proteins are secreted through cellular pathways that do not involve a signal peptide, ProteINSIDE has predicted proteins that may be secreted either by endosomal recycling, plasma membrane transporter, membrane flip-flop, and membrane blebbing that involves formation of vesicles or exosomes. As the results of the 3 methods of prediction (UniProt subcellular location, a GO term related to a secretory function, and TargetP prediction), ProteINSIDE proposed 8 and 7 additional proteins potentially secreted by the muscle and the AT, respectively. Among these additional proteins potentially secreted by the muscle CASQ1 and CASQ2 (Calsequestrin-1 and 2) or TXN (Thioredoxin) were reported to be involved in skeletal muscle development and contraction or in cell proliferation, respectively. The additional proteins potentially secreted by the AT were ANXA7 (Annexin A7) that was reported to regulate cell proliferation and PPIA (Peptidyl-prolyl cis-trans isomerase A) known to positively regulate protein secretion.

Adipose tissue	Muscle tissue	Adipose and muscle tissue
A2M	AHSG	AHSG
ADIPOQ	ALB	ALB
AFM	AMBP	AMBP
AHSG	APOA1	APOA1
ALB	APOH	APOH
AMBP	C1QTNF3	CLEC3B
APOA1	CLEC3B	COL5A1
APOA1BP	COL4A2	COL6A1
APOH	COL5A1	CTSB
C3	COL6A1	CTSD
C4A	CTSB	FGA
CFI	CTSD	FGB
CLEC3B	DPT	GSN
COL1A2	FGA	HPX
COL5A1	FGB	KNG1
COL6A1	GSN	LUM
COL6A3	HPX	OGN
CPB2	KNG1	OLFML3
CRTAP	LAMA2	POSTN
CTSB	LUM	PRDX4
CTSD	OGN	RBP4
DCN	OLFML3	SERPINA1
FETUB	POSTN	SERPINF1
FGA	PRDX4	TTR
FGB	RBP4	
FGG	SERPINA1	
GC	SERPINF1	
GGH	TTR	
GPX3		
GSN		
HPX		
IGJ		
ITIH3		
KNG1		
LUM		
MFAP4		
OGN		
OLFML3		
ORM1		
P4HB		
PCOLCE		
PLG		
POSTN		
PRDX4		
RBP4		
SERPINA1		
SERPINA3		
SERPINA5		
SERPINC1		
SERPINF1		
SERPINF2		
SERPINF1		
TGFBR3		
TTR		

Table 1: Proteins potentially secreted thanks to a signal peptide by bovine foetal AT, muscle or both tissues. The predictions were confirmed by the subcellular location provided by TargetP and Uniprot, and by GO annotations related to the secretory function.

1.3. Proteins interactions

PPI were searched between proteins within (core network) and outside (extended network) the dataset for proteins identified either in muscle or in AT. For the core muscular and adipose networks, the number of PPI identified by querying 3 major databases (UniProt, BioGrid, and IntAct preselected in the basic analysis) were 572 PPI between 271 proteins identified in muscle (over the 490) and 966 PPI between 396 proteins identified in AT (over 705). To identify key proteins for the regulation of fetal tissue growth we have sorted networks by applying high values of betweenness (BC) and closeness (CC) centralities (data not shown) and the same central proteins were retrieved in the muscular and adipose core networks, as the results of the 331 proteins identified both in AT and muscle. The core network for the 331 common proteins was sorted and this has revealed 25 proteins as being centrals in the networks (Supp Figure 1; settings of the network: BC 550 and CC: 0.15). The most central proteins were ALB, CRK, HSPB1, and GRB2 described in context spanning stem cell differentiation, regulation of cell differentiation or apoptosis. Within the extended network, 300 over the 331 common proteins were among the 3415 proteins linked by 8960 PPI. 28 proteins were proposed as being central within the network (Figure 3; settings of the network: BC 120000 and CC: 0.2). The most central proteins of the network were VCAM1, GRB2, HSPB1, and PPP1CA, three of them were proteins identified both in fetal AT and muscle. GO term annotations for the 28 central proteins have shown that these proteins may be involved in the positive regulation of cell proliferation (EGFR and FN1), in cell differentiation (GRB2, FN1), in the autophagic process (GABARAPL1, GABARAPL2, and MAP1LC3A) and in the negative regulation of apoptotic process (HSPB1, HNRNPK, HSPA5, EGFR, PSMA1).

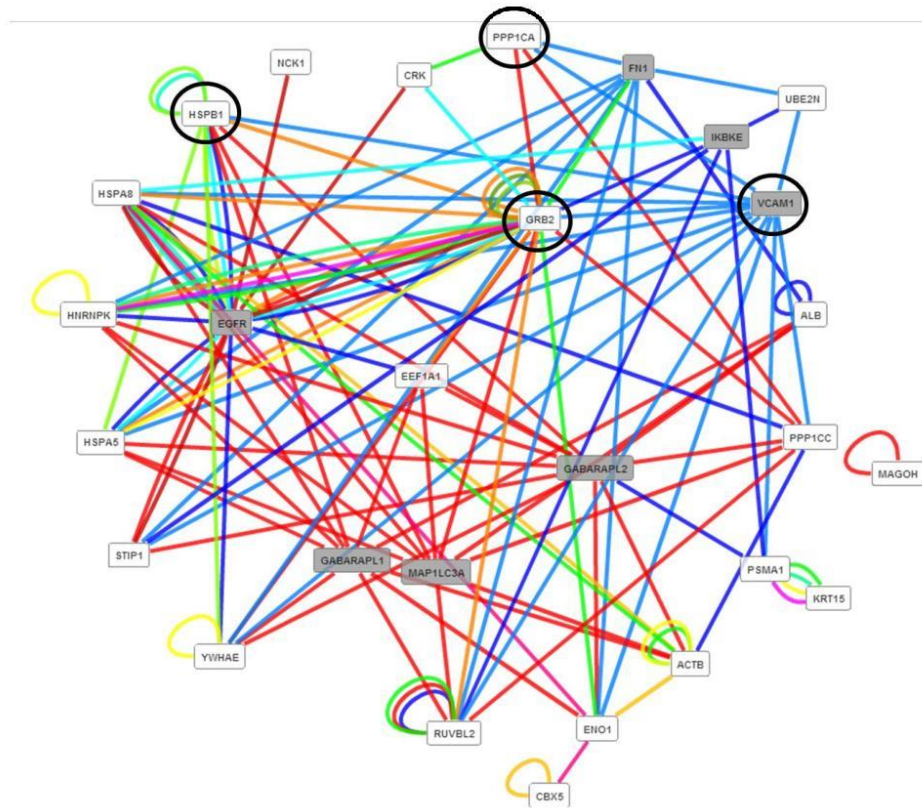


Figure 3: Extended PPI network between the 331 proteins identified both in fetal AT and muscles and external proteins. The network has been created using the PPI from BioGrid, UniProt, and IntAct databases. Squares (white for protein from the dataset, and grey for protein that were not in the dataset) indicate key proteins identified by sorting the network with algorithms of betweenness (an average of 120000) and closeness centralities (an average of 0,2).

The core network for the 173 proteins identified specifically within of the muscle tissue displayed 82 PPI made by 54 protein of the dataset (Supplemental Figure 2). The most central proteins of this networks were MYC (Myc proto-oncogene protein), TTN (Titin), FHL1 (Four and a half LIM domains protein 1), and CCT8 (T-complex protein 1 subunit theta). Within the extended network,

153 over the 173 muscular proteins were among the 1761 proteins linked by 3200 PPI. 17 proteins were proposed as being central within the network (Figure 4; settings of the network: BC 15000 and CC: 0.3). The most central proteins of the network were STUB1 (E3 ubiquitin-protein ligase CHIP), MYC, TTN, and PUF60 (Poly(U)-binding-splicing factor PUF60). GO term annotations for the 17 central proteins have shown that these proteins may be involved in positive regulation of canonical Wnt signaling pathway (PSMC6, LRRK2, and PSMA7), negative regulation of apoptotic process (MYC, PSMA7, BAG3, and PSMC6), muscle contraction (MYL6 and TTN), and autophagy process (LRRK2 and GABARAPL2).

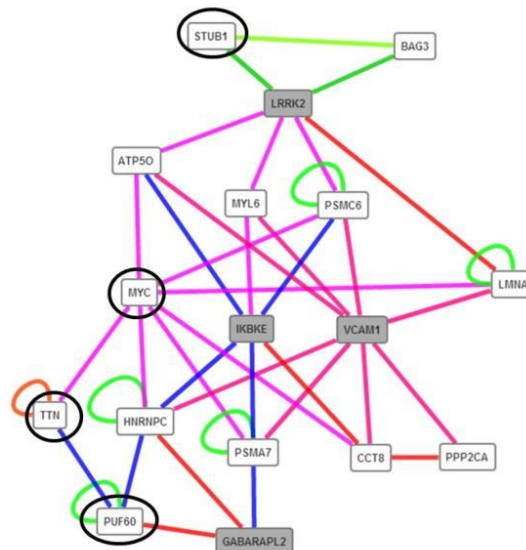


Figure 4: Central proteins within the extended PPI network involving 173 proteins identified only in the fetal muscle tissue. The network has been created using the PPI from BioGrid, UniProt, and IntAct databases. Squares (white for protein from the dataset, and grey for protein that were not in the dataset) indicate key proteins identified by sorting the network with algorithms of betweenness (an average of 15000) and closeness centralities (an average of 0,3).

The core network for the 391 proteins specifically identified in AT displayed 304 PPI made by 185 proteins of the dataset (Supplemental Figure 3). The most central proteins of this network were ARRB1 (Beta-arrestin-1), IMMT (MICOS complex subunit MIC60), CDC42 (Cell division control protein 42 homolog), and TUBA1C (Tubulin alpha-1C chain). Within the extended network, 349 of the 391 adipose proteins were among the 3364 proteins linked by 7692 PPI. 34 proteins were proposed as being central within the network (Figure 5; settings of the network: BC 100000 and CC: 0.32). The most central proteins of the network were VCAM1 (Vascular cell adhesion protein 1), GRB2 (Growth factor receptor-bound protein 2), FLNA (Filamin-A), and YWHAG (14-3-3 protein gamma). GO term annotations for the 17 central proteins have shown that these proteins may be involved in negative regulation of apoptotic process (PSMB1, DDX3X, PSMA3, TRAF6, KRT18, MYC, and FLNA), in positive regulation of canonical Wnt signaling pathway (LRRK2, PSMB1, and PSMA3), in autophagy process (LRRK2, HSP90AA1, and GABARAPL1), and in cell cycle (CSNK2B, NUDC, PSMB1, HSP90AA1, TUBB4A, YWHAG, PSMA3, and TUBG1).

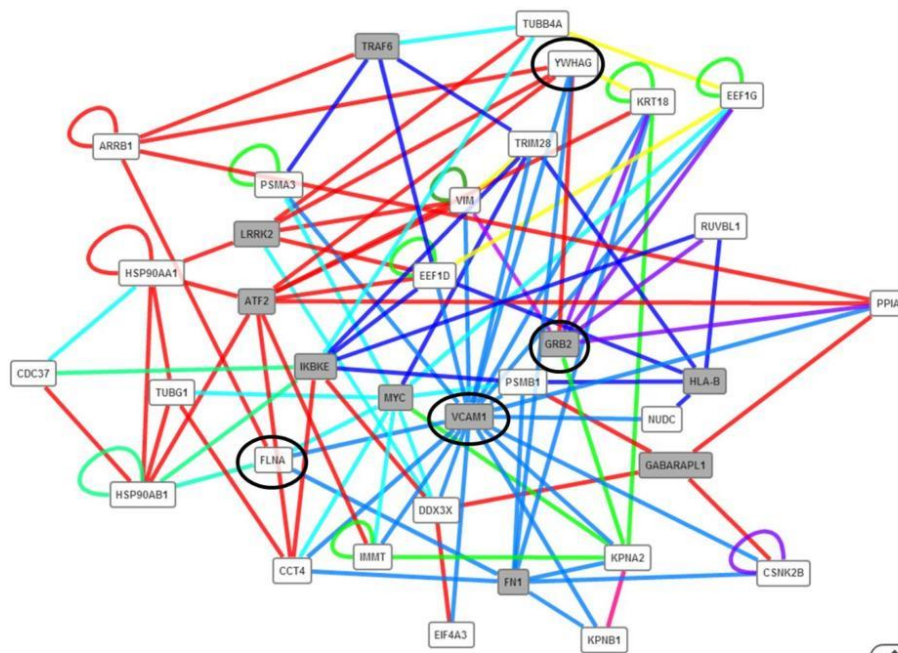


Figure 5: Central proteins within the extended PPI network involving 349 proteins identified only in fetal AT (extended PPI network). The network has been created using the PPI from BioGrid, UniProt, and IntAct databases. Squares (white for protein from the dataset, and grey for protein that were not in the dataset) indicate key proteins identified by sorting the network with algorithms of betweenness (an average of 100000) and closeness centralities (an average of 0,32).

III. Highlighting of proteins interactions between adipose and muscles tissues

In order to understand the “adipose-muscular” cross-talk we have searched for known interactions between proteins secreted by one tissue and proteins present at the cell surface of the other tissue.

III.1. Proteins secreted by AT that interact with cell surface proteins of the muscle

We have created a network between the 67 proteins potentially secreted by the bovine fetal AT and the 55 cell surface proteins that we have identified among the muscular proteins (Figure 6). 32 links involved in regulatory networks were identified and they linked 29 proteins. Among these regulatory links two have been highlighted and analysed using Pathway studio. The first PPI (A) engaged the protein CLEC3B (tetranectin), synthesized and secreted by the AT, with PLG (Plasminogen) and ENO1 identified in the proteome of the fetal muscle. The second couple (B) engaged the protein DCN (Decorin), synthesized and secreted by the AT, and THBS1 (Thrombospondin-1) identified in the proteome of the fetal muscle.

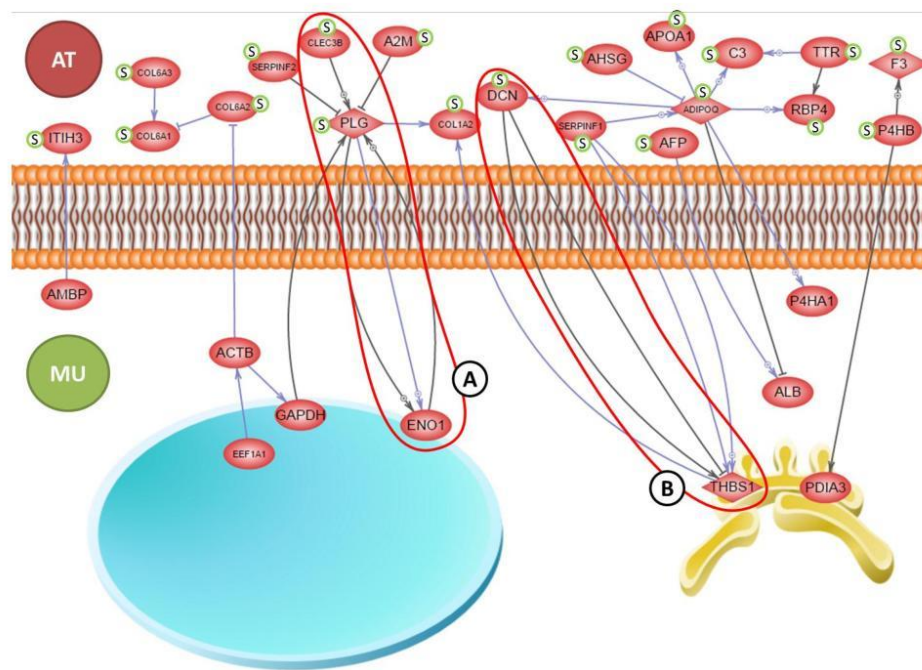


Figure 6: Network of interacting proteins between the proteins potentially secreted by AT and the cell surface proteins of the muscle (MU). Proteins potentially secreted by AT were highlighted by an

“S” within a green circle and those potentially secreted both by AT. We also highlighted the two ways of interactions between each tissues (A and B). Blue arrow are interactions for “regulation the expression” and black arrow are “direct physical” interactions as defined by Pathway studio.

III.2. Proteins secreted by the muscle that interact with cell surface proteins of AT

We have created a network between the 33 proteins potentially secreted by the bovine fetal muscle and the 106 cell surface proteins that we have identified among the adipose proteins (Figure 7). We have identified 32 links involved in regulatory networks between 32 proteins. We highlighted two PPI, the first (A) was between the protein C1QTNF3 (Complement C1q tumor necrosis factor-related protein 3), secreted and synthesized by the muscle, and the protein ADIPOQ (Adiponectin), synthesized by the AT; the second couple (B) between the protein DPT (Dermatopontin), secreted and synthesized by the muscle, and the protein DCN.

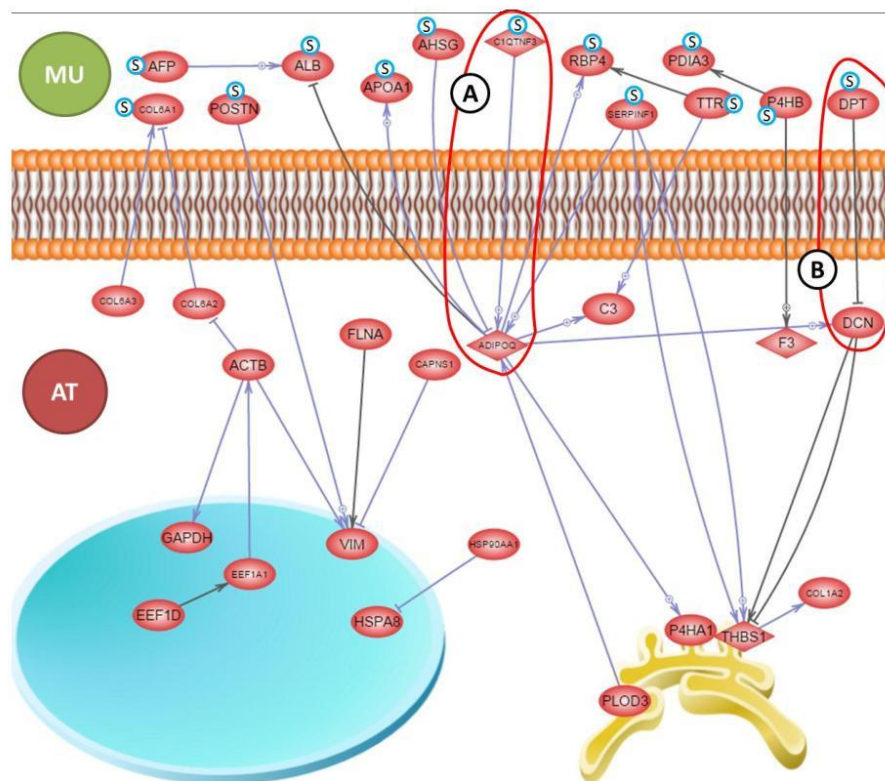


Figure 7: network of interacting proteins between the potentially secreted proteins of the muscle and the cell surface proteins of the AT. We highlighted the potentially secreted proteins of the muscle with an “S” on a blue circle. We also highlighted the two ways of interactions between each tissues (A and B). Blue arrow are interaction for “regulation the expression” and black arrow are “direct physical” interaction.

DISCUSSION

This study is the first one which established an enriched proteome of muscle and adipose tissue for the same fetuses. It provides an accurate description of the main functional pathways and

biological functions involved in myogenesis, adipogenesis, and in the cross-talk between the two tissues. The most original results concern the prediction of secreted proteins and their potential role in the regulation of the lean-to-fat ratio or in the cross-talk between AT-muscle.

With the present proteome, we identified 607 and 411 new proteins respectively for the fetal AT and skeletal muscle in bovine species. In agreement with our knowledge of myogenesis and adipogenesis, we showed that the main terms GO concern the proliferation and differentiation of both tissues. In 140 days-old bovine fetuses, myogenesis is characterized by the proliferation of the secondary and tertiary generations of myoblasts and by the differentiation of the myotubes mainly from the primary generation (for review: (Picard *et al.*, 2010; Picard *et al.*, 2002)). At this age in the present study we have identified many GO terms corresponding to structural organisation of muscle fibers and contractile differentiation. This is consistent with the events occurring at this age corresponding to the synthesis of myofibrillar proteins and the ultrastructural organization inside myofibres well described *in vitro* in muscle cell culture or *in vivo* in bovine and other species (Picard *et al.*, 2010). In the muscle PPI network, TTN (Titin) protein appeared with several interactions with other proteins of our muscular proteome. This “giant” protein plays a central role in striated muscle structure and contraction (Labeit et Kolmerer, 1995). It is considered as a good marker of muscle differentiation in muscle cell cultures (vanderLoop *et al.*, 1996).

On 140 dpc, the pre-adipocyte cells proliferate and differentiate in adipocytes (for review: (Bonnet *et al.*, 2010)). The lipid metabolism is activated in new formed adipocytes in agreement with the GO terms corresponding to lipid metabolism revealed in the present study. The AT PPI network showed core position of ActB, ARRB1 and Hsp90AB1. The identification of ActB as a key protein is consistent with the cytoskeletal remodeling necessary for adipogenesis (Gregoire *et al.*, 1998) as exemplified by strong variations in the abundance of most of the cytoskeletal proteins, including vimentin, tubulin, actins and tropomyosin, during the differentiation of 3T3-L1 fibroblasts into adipocytes (Molina *et al.*, 2009; Welsh *et al.*, 2004). In the context of AT biology, ARRB1 was only

reported as a Golgi apparatus gene expressed in extracellular vesicles from porcine adipose tissue-derived mesenchymal stem cells (Eirin *et al.*, 2014). Hsp90AB1 is a molecular chaperone that promotes the maturation, structural maintenance and proper regulation of specific target proteins involved for instance in cell cycle control and signal transduction (Uniprot) and has been rarely identified in AT. The expression of Hsp90AB1 has been reported in differentiating 3T3-L1 preadipocytes as a chaperone that controls the misfolding of nascent poly-peptide chains (Fromm-Dornieden *et al.*, 2012). Hsp90AB1 has also been proposed as a cell surface markers of of multipotential mesenchymal precursor cells (Gronthos *et al.*, 2009).

The analysis of the PPI networks showed that 3 proteins, HSPB1, GRB2 and VCAM1, have a core position in the two tissues. HSPB1, a member of small Hsp (sHsp) family, is expressed in all human tissues, mainly in skeletal, smooth, and cardiac muscles (Sugiyama *et al.*, 2000). Expression of HSPB1 is induced during ontogenesis (Lutsch *et al.*, 1997), in various pathologies and in many types of tumors (Ciocca et Calderwood, 2005). HSPB1 has several modes of action, including limiting the number of misfold proteins product during a stress event like cell differentiation, interacting and stimulating Akt protein preserving the cell from death and HSPB1 can bind directly in the caspase cascade from a mitochondrial/ CytC bax mediated signal to inhibit Caspase 3 activity (Bryantsev *et al.*, 2007). Other roles in preventing cells from apoptosis through actin-cytoskeleton remodelling were also described (Bruneel *et al.*, 2005). A role of HSPB1 in the protection against oxidative stress has also been described (Kammoun *et al.*, 2013). Chaze et al (Chaze *et al.*, 2008), in bovine muscle, showed an upregulation of HSPB1 during the first two thirds of foetal life concomitant with an increased control of cell death through this time. Similar variations in the abundance of HSPB1 were reported in the fetal AT (Taga *et al.*, 2012). Serum level of HSPB1 was associated with the level of adiposity in human (Tavallaie *et al.*, 2011) and to the lipid metabolism in bovine AT (Rajesh *et al.*, 2010). The protein GRB2 (Growth factor receptor-bound protein 2) is an adapter protein that provides a critical link between cell surface growth factor receptors and the downstream signaling pathways. In murine muscle, GRB2 is part of the Hepatocyte Growth factor signaling required for the

developpement of secondary fibers (Maina *et al.*, 1996). In fetal murine progenitors cells GRB2 can be a downstream effector of Hepatocyte Growth factor (Maina *et al.*, 1996). In AT, GRB2 is a downstream effector of the insulin-like growth factor 1 signaling that is known to bifurcate to lead to either proliferation or differentiation of the adipose progenitors (Woldt *et al.*, 2011). The identification of GRB2 as central protein both for AT and muscle in the present study is consistent with the contribution of the second generation of fibers to muscle growth at 140 dpc, and to the balance between proliferation and differentiation for AT growth. The identification of GRB2 partners in the present proteomes may help to identify regulators of these pathways for tissues growth in bovine. VCAM1 also named CD106 is a cell surface marker of mesenchymal stem cells was already observed as a central protein in our previous analysis (Kaspric *et al.*, 2015) and should indicate that mesenchymal stem cells contribute to the post-embryonic growth of AT and muscle as already proposed.

Present results reinforced our previous hypothesis of an engagement of autophagic proteins to the processes of adipogenesis and myogenesis (Kaspric *et al.*, 2015). The present study complement our previous data by the identification of new proteins that were described in the context of the autophagy process (CTSD, HMGB1, and ANXA7; (Li *et al.*, 2013; Tang *et al.*, 2010)), especially in its positive regulation (PAFAH1B2 and SH3GLB1; (McKnight *et al.*, 2012; Takahashi *et al.*, 2007)) and the formation of autophagosome necessary for the macroautophagy process (RAB1A; (Dong *et al.*, 2012)).

Proteins related to the Wnt pathway have been identified both in AT and muscle, in agreement with the known importance of this pathway for the balance between myogenesis and adipogenesis. Indeed, activation of Wnt signaling stabilizes β -catenin, which regulates the expression of myogenic (pax 3, MyoD) and adipogenic (PPAR- γ) transcription factors (Du *et al.*, 2011). Over-expression of β -catenin results in increased myoblast proliferation (Otto *et al.*, 2008) and a repressed

adipogenesis. The present study has revealed proteins never identified in AT and muscle such as TBL1XR1 as potential targets or regulatory partners of the Wnt signaling.

Another original result of the present study is the revelation of many GO terms related to apoptotic processes. In the two tissues these GO terms were the most numerous. Among them the GO terms corresponding to a negative regulation of apoptosis were the most represented process. These data are in agreement with the results of Chaze et al (Chaze *et al.*, 2008) in the *Semitendinosus* muscle of bovine suggesting an important role of apoptosis. These authors revealed a fine balance between cell proliferation and apoptosis during the two first trimesters of gestation in bovine muscle, leading to the control of the total number of fibres. This effect of massive negative regulation could be necessary to prioritize cell proliferation as already view in guinea pigs (Sun et al., 2014). The present study has revealed 19 muscular and 31 adipose news proteins by comparison with the previous fetal proteome (Chaze *et al.*, 2008; Chaze *et al.*, 2009; Taga *et al.*, 2012). For both tissues, the newly identified proteins are mainly subunits of proteasome which are more numerous in AT than muscle (15 proteins in AT and 9 in the muscle). This may be related to the involvement of the proteasome in the inhibition of differentiation of the adipose progenitors by the degradation of ubiquitinated adipogenic transcription factors PPAR- γ (Zhang *et al.*, 2013) et CEBP- α (Pal *et al.*, 2013). A fine balance between autophagy and ubiquitination could be determinant for the balance between proliferation and differentiation by modulating the abundance of the master gene of adipogenesis PPAR- γ .

The other proteins related to negative regulation of apoptosis identified in the present study in the two tissues are GLO1, HSPA5, THSBS1, TPT1. The protein GLO1 (Lactoylglutathione lyase or Glyoxalase I) is expressed in several tissues. The enzyme encoded by this gene is responsible for the catalysis and formation of S-lactoyl-glutathione from methylglyoxal condensation and reduced glutathione. In osteoblasts it was described to be involved in the regulation of TNF-induced transcriptional activity of NF-kappa-B. Mahmoud *et al* (Mahmoud *et al.*, 2013), in human skeletal

muscle, proposed that GLO1 may play an important role in the maintenance of healthy skeletal muscle metabolism and that aerobic exercise training is effective in normalizing GLO1 protein expression in obese insulin resistant subjects. Its role in adipogenesis or myogenesis has never been described. The Tumor suppressor p53 (TPT1) maintains genome stability by differentially activating target genes controlling cellular processes such as the antioxidant response, cell cycle arrest and apoptosis. TPT1 is a direct target gene of p53 known to induce the intrinsic pathway of apoptosis. p53-dependent induction of TPT1 was shown to be able to reduce oxidative stress, minimize apoptosis, and promote cell survival in response to H₂O₂ challenge in lung tissues (Chen *et al.*, 2013). HSPA5 found in the two tissues as a negative regulator of apoptosis, encodes a member of the heat shock protein 70 (HSP70) family. It is localized in the lumen of the endoplasmic reticulum. The members of the Hsp70 family serve a variety of roles including folding of nascent polypeptides, protein translocation across membranes, chaperone-mediated autophagy, and prevention of protein aggregation under stress conditions (Daugaard *et al.*, 2007). Several studies showed that Hsp70 is one of the most important heat shock protein for maintenance of cell integrity during normal cellular growth as well as under pathophysiological conditions. An anti-apoptotic role of Hsp70 has been proposed in skeletal muscle by Gao *et al.* (2010). They reported a function of Hsp70 in the regulation of TNF- α -induced cell apoptosis through a complex Hsp70/CHIP/ASK1 in which Hsp70 promotes ASK1 proteasomal degradation and prevents TNF- α -induced cell apoptosis. Hsp70 proteins are also known to sequester pro-apoptotic factors such as BCL-2 (Daugaard *et al.*, 2007). Other proteins annotated with GO term relative to apoptosis such as BAG3, GOLPH3, HSPD1, LTF, PAK2, UBA52 were found only in skeletal muscle tissue, while CAT, CFL1, FHL2, FLNA, YWHAZ were identified in AT. The analysis of the corresponding biological functions will allow to better understand tissue-specificity in the regulation of apoptosis for the growth of both tissues.

The most original contribution of the present study is the identification of secreted proteins which constitute good candidates for the interactions between the two tissues and the cross-talk between AT and muscle. We confirmed that ADIPOQ, AFP, ERLIN2, FGG, NDUFA10, PCCB, PDIA3,

RCN1, SERPINA1, TTR, ALB, APOA1, GSN, and SERPINA1, could be secreted both by AT and muscle as in our previous analyses (Kaspric *et al.*, 2015). Our study brought 24 new proteins potentially secreted by the muscle. Some of these proteins may contribute to the pathways that we have discussed to sustain myogenesis: regulation of cell proliferation (OGN (Mimcan, UniProtKB and (Arita *et al.*, 2002), RBP4 (Retinol-binding protein 4), MAP2K1 (Dual specificity mitogen-activated protein kinase kinase 1; (Lin *et al.*, 1998)), regulation of apoptosis (APOH (Lin *et al.*, 2005) and KNG1 (Kininogen-1; (Al-Fakhri *et al.*, 2003))) and LGALS1 (Galectin-1; (Bruneel *et al.*, 2005)), or muscle development (LAMA2, Laminin subunit alpha-2; (Helbling-leclerc *et al.*, 1995)). Our study brought 48 new proteins potentially secreted by the fetal AT. They could be involved in the regulation of apoptosis (PLG protein (Plasminogen; (Rossignol *et al.*, 2004)), lipid metabolism (C3 (Complement C3; (Saleh *et al.*, 1998)), autophagy (DCN (Buraschi *et al.*, 2013), RAB1A (Mukhopadhyay *et al.*, 2011)), or cell differentiation (SERPINF2 (Alpha-2-antiplasmin; (Mukhopadhyay *et al.*, 2011))). These proteins may also contribute to the “adipose-muscular” cross-talk .

The networks of interacting proteins between the potentially secreted proteins of one tissue and the cell surface proteins of the other tissue, have highlighted two ways of interactions for both tissues: direct physical interaction or interaction through a regulatory networks. Among the proteins potentially secreted by AT and having a potential role in the regulation of muscle, we have proposed CLEC3B and DCN. CLEC3B (Tetranectin) is known to upregulate ENO1 (Alpha-enolase; (Diaz-Ramos *et al.*, 2012)) within muscle through a regulatory network that involves PLG synthesised by the liver but identified in the fetal muscle. Interestingly ENO1 was identified in muscle early in the fetal bovine life and was proposed to favor the proliferation (Chaze *et al.*, 2008) because of its known inhibition of the cellular growth (Ghosh *et al.*, 1999). DCN is a secreted proteoglycan, it is one of the most studied members of the small leucine-rich proteoglycan family. It has been identified in various tissues including skeletal muscle (for review: (Nishimura *et al.*, 2015)). Circulating DCN was shown to be elevated in the plasma of diabetic (Bolton *et al.*, 2008). In muscle and AT, DCN interacts with THBS1 (synthesized by each tissues) to block its expression. THBS1 is an adhesive glycoprotein that mediates

cell-to-cell and cell-to-matrix interactions by binding to fibrinogen, fibronectin, laminin, type V collagen and integrins alpha-V/beta-1. It is known to be involved in the negative regulation of apoptotic processes (Pallero *et al.*, 2008) that we have proposed as a key biological function both for fetal AT and muscle growth.

The muscular proteins that could modulate the fetal adipose tissue growth as suggested by the present results were C1QTNF3 and DPT. C1QTNF3 is known to be involved in the positive regulation of the ADIPOQ secretion (Wolfing *et al.*, 2008) and the regulation of the cell fat differentiation (Kopp *et al.*, 2010). ADIPOQ is known to be involved in the negative regulation of the fat cell differentiation (Yokota *et al.*, 2002) and of the muscle cell proliferation (Arita *et al.*, 2002). The Adiponectin is also known to be an important adipokine involved in the control of fat metabolism and insulin sensitivity (Yamauchi *et al.*, 2002). Its expression is regulated by several proteins potentially secreted by the muscle. DPT, a low-molecular-mass component of the extracellular matrix, is known to be involved in the negative regulation of cell proliferation (Tzen et Huang, 2004). Okamoto *et al.* (Okamoto *et al.*, 1999) showed in endothelial cells that DPT inhibits the formation of the DCN/TGF- β 1 complex. However, its role in skeletal muscle or adipose tissue has not been described. The available knowledge concerns the DCN/TGF- β 1 pathway in muscle. It was demonstrated that DCN binds to TGF- β through its core-protein and sequesters TGF- β (inhibitor of both proliferation and differentiation of myogenic cells) by trapping it in the extracellular matrix. Some results on C2C12 line, demonstrated that DCN stimulates TGF- β dependent signaling via lipoprotein-receptor related protein (LRP-1) in non differentiated myoblasts. In differentiated myotubes expressing low level of LRP-1, DCN inhibits signaling by sequestering TGF- β . This suggests that DCN has differential effects on TGF- β dependent signaling at the early and late stages of myogenesis. These data are in favour of a role of DCN, and probably by this way of DPT, in the cross-talk between muscle and AT. To confirm this, it could be interesting to analyse the expression of DCN throughout fetal development in bovine muscle and AT. In chicken skeletal muscle, it was shown that its expression increases from 14 to 19 days of embryonic growth and decrease prior to hatching.

Kishioka *et al.* (Kishioka *et al.*, 2008) demonstrated that DCN enhances the differentiation of myoblasts by suppressing myostatin (a member of TGF-Beta family) activity. Other authors suggested that DCN acts as a signaling molecule through an activation of Akt downstream of IGF-IR (Suzuki *et al.*, 2013). However, there are no reports of this action on myogenic cells. More recently it was proposed that DCN is involved in the positive regulation of autophagic processes in endothelial cells (Buraschi *et al.*, 2013). It induces Peg3 activating the transcription of autophagic proteins such as Beclin 1 and LC3. These data demonstrate that messages embedded in the extracellular environment can profoundly affect autophagy. However this role of DCN has to been demonstrated in other cells.

To conclude, this study provides new data on the molecular events underpinning simultaneously bovine myogenesis and adipogenesis at a fetal age corresponding to an increase both in cell number and volume for AT and muscle. These data are valuable for issues related to the lean-to-fat ratios of bovine carcasses and meat for the beef industry, but also for knowledge relevant to human body composition ontogenesis. Indeed, available data revealed that myogenesis in cattle is very close to that of human (same types of fibres, same gestation length, similar kinetics of expression of developmental Myosin heavy chain isoforms, total number of fibres set about at the end of second trimester ...).

ABBREVIATIONS

AT: Adipose Tissue

Da: Dalton

dpc: day post conception

GO: Gene Ontology

ID: Identifier

min: minute

mm: millimeter

PPI: Protein-Protein interaction

%: percent / percentage

µm: micrometers

REFERENCES

- Al-Fakhri, N., Chavakis, T., Schmidt-Woll, T., Huang, B., Cherian, S.M., Bobryshev, Y.V., Lord, R.S., Katz, N. and Preissner, K.T. Induction of apoptosis in vascular cells by plasminogen activator inhibitor-1 and high molecular weight kininogen correlates with their anti-adhesive properties. *Biol. Chem.* 2003;384(3):423-435.
- Argiles, J.M., Orpi, M., Busquets, S. and Lopez-Soriano, F.J. Myostatin: more than just a regulator of muscle mass. *Drug Discov. Today* 2012;17(13-14):702-709.
- Arita, Y., Kihara, S., Ouchi, N., Maeda, K., Kuriyama, H., Okamoto, Y., Kumada, M., Hotta, K., Nishida, M., Takahashi, M., Nakamura, T., Shimomura, I., Muraguchi, M., Ohmoto, Y., Funahashi, T. and Matsuzawa, Y. Adipocyte-derived plasma protein adiponectin acts as a platelet-derived growth factor-BB-binding protein and regulates growth factor-induced common postreceptor signal in vascular smooth muscle cell. *Circulation* 2002;105(24):2893-2898.
- Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R.L., Yoon, C., Schiess, R., Schmidt, A., Mirkowska, P., Hartlova, A., Van Eyk, J.E., Bourquin, J.P., Aebersold, R., Boheler, K.R., Zandstra, P. and Wollscheid, B. A mass spectrometric-derived cell surface protein atlas. *PLoS one* 2015;10(3):e0121314.
- Bolton, K., Segal, D., McMillan, J., Jowett, J., Heilbronn, L., Abberton, K., Zimmet, P., Chisholm, D., Collier, G. and Walder, K. Decorin is a secreted protein associated with obesity and type 2 diabetes. *Int. J. Obes.* 2008;32(7):1113-1121.
- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal : an international journal of animal bioscience* 2010;4(7):1093-1109.
- Bouley, J., Chambon, C. and Picard, B. Mapping of bovine skeletal muscle proteins using two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* 2004;4(6):1811-1824.
- Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Royer, N., Vinh, J., Pernet, P., Vaubourdolle, M. and Baudin, B. Proteomics of human umbilical vein endothelial cells applied to etoposide-induced apoptosis. *Proteomics* 2005;5(15):3876-3884.
- Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Royer, N., Vinh, J., Pernet, P., Vaubourdolle, M. and Baudin, B. Proteomics of human umbilical vein endothelial cells applied to etoposide-induced apoptosis. *Proteomics* 2005;5(15):3876-3884.
- Bryantsev, A.L., Kurchashova, S.Y., Golyshev, S.A., Polyakov, V.Y., Wunderink, H.F., Kanon, B., Budagova, K.R., Kabakov, A.E. and Kampinga, H.H. Regulation of stress-induced intracellular sorting and chaperone function of Hsp27 (HspB1) in mammalian cells. *Biochem. J.* 2007;407:407-417.
- Buraschi, S., Neill, T., Goyal, A., Poluzzi, C., Smythies, J., Owens, R.T., Schaefer, L., Torres, A. and Iozzo, R.V. Decorin causes autophagy in endothelial cells via Peg3. *Proceedings of the National Academy of Sciences of the United States of America* 2013;110(28):E2582-2591.
- Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguluy, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K. and Tyers, M. The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 2015;43(Database issue):D470-478.
- Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B. In vivo proteome dynamics during early bovine myogenesis. *Proteomics* 2008;8(20):4236-4248.
- Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B. Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle. *Animal : an international journal of animal bioscience* 2009;3(7):980-1000.
- Chen, W.M., Wang, H.H., Tao, S.S., Zheng, Y., Wu, W., Lian, F.R., Jaramillo, M., Fang, D.Y. and Zhang, D.D. Tumor protein translationally controlled 1 is a p53 target gene that promotes cell survival. *Cell Cycle* 2013;12(14):2321-2328.

- Ciocca, D.R. and Calderwood, S.K. Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress Chaperones* 2005;10(2):86-103.
- Coordinators, N.R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2015;43(Database issue):D6-17.
- Daugaard, M., Kirkegaard-Sorensen, T., Ostefeld, M.S., Aaboe, M., Hoyer-Hansen, M., Orntoft, T.F., Rohde, M. and Jaattela, M. Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer. *Cancer Res.* 2007;67(6):2559-2567.
- Dong, N., Zhu, Y., Lu, Q., Hu, L., Zheng, Y. and Shao, F. Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell* 2012;150(5):1029-1041.
- Draeger, A., Weeds, A.G. and Fitzsimons, R.B. Primary, Secondary and Tertiary Myotubes in Developing Skeletal-Muscle - a New Approach to the Analysis of Human Myogenesis. *J. Neurol. Sci.* 1987;81(1):19-43.
- Du, M., Zhao, J.X., Yan, X., Huang, Y., Nicodemus, L.V., Yue, W., McCormick, R.J. and Zhu, M.J. Fetal muscle development, mesenchymal multipotent cell differentiation, and associated signaling pathways. *J. Anim. Sci.* 2011;89(2):583-590.
- Eirin, A., Riester, S.M., Zhu, X.Y., Tang, H., Evans, J.M., O'Brien, D., van Wijnen, A.J. and Lerman, L.O. MicroRNA and mRNA cargo of extracellular vesicles from porcine adipose tissue-derived mesenchymal stem cells. *Gene* 2014;551(1):55-64.
- Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 2000;300(4):1005-1016.
- Fromm-Dornieden, C., von der Heyde, S., Lytovchenko, O., Salinas-Riester, G., Brenig, B., Beissbarth, T. and Baumgartner, B.G. Novel polysome messages and changes in translational activity appear after induction of adipogenesis in 3T3-L1 cells. *BMC Mol. Biol.* 2012;13.
- Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015;43(Database issue):D1049-1056.
- Ghosh, A.K., Steele, R. and Ray, R.B. Functional domains of c-myc promoter binding protein 1 involved in transcriptional repression and cell growth regulation. *Mol. Cell. Biol.* 1999;19(4):2880-2886.
- Gonzalez, J.M., Camacho, L.E., Ebarb, S.M., Swanson, K.C., Vonnahme, K.A., Stelzleni, A.M. and Johnson, S.E. Realimentation of nutrient restricted pregnant beef cows supports compensatory fetal muscle growth. *Journal of animal science* 2013;91(10):4797-4806.
- Gregoire, F.M., Smas, C.M. and Sul, H.S. Understanding adipocyte differentiation. *Physiol. Rev.* 1998;78(3):783-809.
- Gronthos, S., McCarty, R., Mrozik, K., Fitter, S., Paton, S., Menicanin, D., Itescu, S., Bartold, P.M., Xian, C. and Zannettino, A.C.W. Heat Shock Protein-90 beta Is Expressed at the Surface of Multipotential Mesenchymal Precursor Cells: Generation of a Novel Monoclonal Antibody, STRO-4, With Specificity for Mesenchymal Precursor Cells From Human and Ovine Tissues. *Stem Cells Dev.* 2009;18(9):1253-1261.
- Gu, B., Zhang, J., Wang, W., Mo, L., Zhou, Y., Chen, L., Liu, Y. and Zhang, M. Global expression of cell surface proteins in embryonic stem cells. *PLoS one* 2010;5(12):e15795.
- Hausman, G.J., Basu, U., Wei, S., Hausman, D.B. and Dodson, M.V. Preadipocyte and Adipose Tissue Differentiation in Meat Animals: Influence of Species and Anatomical Location. *Annu Rev Anim Biosci* 2014;2:323-351.
- Helbling-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., Weissenbach, J., Tome, F.M.S., Schwartz, K., Fardeau, M., Tryggvason, K. and Guicheney, P. Mutations in the Laminin Alpha-2-Chain Gene (Lama2) Cause Merosin-Deficient Congenital Muscular-Dystrophy. *Nat. Genet.* 1995;11(2):216-218.
- Kammoun, M., Picard, B., Henry-Berger, J. and Cassar-Malek, I. A network-based approach for predicting Hsp27 knock-out targets in mouse skeletal muscles. *Computational and Structural Biotechnology Journal* 2013;6(7).

- Kaspric, N., Picard, B., Reichstadt, M., Tournayre, J. and Bonnet, M. ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. *PLoS one* 2015;10(5):e0128086.
- Kaspric, N., Reichstadt, M., Picard, B., Tournayre, J. and Bonnet, M. Protein Function Easily Investigated by Genomics Data Mining Using the ProteINSIDE Online Tool. *Genomics and Computational Biology* 2015;1(1):e16.
- Kishioka, Y., Thomas, M., Wakamatsu, J.I., Hattori, A., Sharma, M., Kambadur, R. and Nishimura, T. Decorin enhances the proliferation and differentiation of myogenic cells through suppressing myostatin activity. *J. Cell. Physiol.* 2008;215(3):856-867.
- Kopp, A., Bala, M., Buechler, C., Falk, W., Gross, P., Neumeier, M., Scholmerich, J. and Schaffler, A. C1q/TNF-related protein-3 represents a novel and endogenous lipopolysaccharide antagonist of the adipose tissue. *Endocrinology* 2010;151(11):5267-5278.
- Labeit, S. and Kolmerer, B. Titins - Giant Proteins in Charge of Muscle Ultrastructure and Elasticity. *Science* 1995;270(5234):293-296.
- Landis, M.D., Carstens, G.E., McPhail, E.G., Randel, R.D., Green, K.K., Slay, L. and Smith, S.B. Ontogenic development of brown adipose tissue in Angus and Brahman fetal calves. *Journal of animal science* 2002;80(3):591-601.
- Lehnert, S.A., Reverter, A., Byrne, K.A., Wang, Y., Natrass, G.S., Hudson, N.J. and Greenwood, P.L. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev. Biol.* 2007;7.
- Li, H., Huang, S., Wang, S., Wang, L., Qi, L., Zhang, Y., Zhang, S., Zhao, B. and Miao, J. Relationship between annexin A7 and integrin beta4 in autophagy. *Int. J. Biochem. Cell Biol.* 2013;45(11):2605-2611.
- Lin, A.W., Barradas, M., Stone, J.C., van Aelst, L., Serrano, M. and Lowe, S.W. Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. *Genes Dev.* 1998;12(19):3008-3019.
- Lin, K.Y., Wang, H.H., Lai, S.T., Pan, J.P. and Chiang, A.N. beta(2)-glycoprotein I protects J774A.1 macrophages and human coronary artery smooth muscle cells against apoptosis. *J. Cell. Biochem.* 2005;94(3):485-496.
- Lutsch, G., Vetter, R., Offhaus, U., Wieske, M., Grone, H.J., Klemenz, R., Schimke, I., Stahl, J. and Benndorf, R. Abundance and location of the small heat shock proteins HSP25 and alpha B-crystallin in rat and human heart. *Circulation* 1997;96(10):3466-3476.
- Mahmoud, A.M., Mey, J.T., Solomon, T.P.J., Fealy, C.E., Malin, S.K., Kirwan, J.P. and Haus, J.M. Glyoxalase 1 protein expression in human skeletal muscle is normalized following aerobic exercise training in obese insulin resistant adults. 2013.
- Maina, F., Casagrande, F., Audero, E., Simeone, A., Comoglio, P.M., Klein, R. and Ponzetto, C. Uncoupling of Grb2 from the Met receptor in vivo reveals complex roles in muscle development. *Cell* 1996;87(3):531-542.
- McKnight, N.C., Jefferies, H.B.J., Alemu, E.A., Saunders, R.E., Howell, M., Johansen, T. and Tooze, S.A. Genome-wide siRNA screen reveals amino acid starvation-induced autophagy requires SCOC and WAC. *EMBO J.* 2012;31(8):1931-1946.
- Molina, H., Yang, Y., Ruch, T., Kim, J.W., Mortensen, P., Otto, T., Nalli, A., Tang, Q.Q., Lane, M.D., Chaerkady, R. and Pandey, A. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *J. Proteome Res.* 2009;8(1):48-58.
- Mukhopadhyay, A., Nieves, E., Che, F.Y., Wang, J., Jin, L., Murray, J.W., Gordon, K., Angeletti, R.H. and Wolkoff, A.W. Proteomic analysis of endocytic vesicles: Rab1a regulates motility of early endocytic vesicles. *J. Cell Sci.* 2011;124(Pt 5):765-775.
- Nikitin, A., Egorov, S., Daraselia, N. and Mazo, I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* 2003;19(16):2155-2157.
- Nishimura, D., Sakai, H., Sato, T., Sato, F., Nishimura, S., Toyama-Sorimachi, N., Bartsch, J.W. and Sehara-Fujisawa, A. Roles of ADAM8 in elimination of injured muscle fibers prior to skeletal muscle regeneration. *Mech. Dev.* 2015;135:58-67.

- Okamoto, O., Fujiwara, S., Abe, M. and Sato, Y. Dermatopontin interacts with transforming growth factor beta and enhances its biological activity. *Biochem. J.* 1999;337:537-541.
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G. and Hermjakob, H. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 2014;42(Database issue):D358-363.
- Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkman, F.S., Cesareni, G., Chatr-aryamontri, A., Chautard, E., Chen, C., Dumousseau, M., Goll, J., Hancock, R.E., Hannick, L.I., Jurisica, I., Khadake, J., Lynn, D.J., Mahadevan, U., Perfetto, L., Raghunath, A., Ricard-Blum, S., Roechert, B., Salwinski, L., Stumpflen, V., Tyers, M., Uetz, P., Xenarios, I. and Hermjakob, H. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nature methods* 2012;9(4):345-350.
- Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stumpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., Salama, J.J., Moore, S., Wojcik, J., Bader, G.D., Vidal, M., Cusick, M.E., Gerstein, M., Gavin, A.C., Superti-Furga, G., Greenblatt, J., Bader, J., Uetz, P., Tyers, M., Legrain, P., Fields, S., Mulder, N., Gilson, M., Niepmann, M., Burgoon, L., De Las Rivas, J., Prieto, C., Perreau, V.M., Hogue, C., Mewes, H.W., Apweiler, R., Xenarios, I., Eisenberg, D., Cesareni, G. and Hermjakob, H. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nature biotechnology* 2007;25(8):894-898.
- Otto, A., Schmidt, C., Luke, G., Allen, S., Valasek, P., Muntoni, F., Lawrence-Watt, D. and Patel, K. Canonical Wnt signalling induces satellite-cell proliferation during adult skeletal muscle regeneration. *J. Cell Sci.* 2008;121(17):2939-2950.
- Pal, P., Lochab, S., Kanaujiya, J.K., Kapoor, I., Sanyal, S., Behre, G. and Trivedi, A.K. E3 Ubiquitin Ligase E6AP Negatively Regulates Adipogenesis by Downregulating Proadipogenic Factor C/EBPalpha. *PLoS one* 2013;8(6).
- Pallero, M.A., Elzie, C.A., Chen, J., Mosher, D.F. and Murphy-Ullrich, J.E. Thrombospondin 1 binding to calreticulin-LRP1 signals resistance to anoikis. *FASEB J.* 2008;22(11):3968-3979.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 2011;8(10):785-786.
- Petryszak, R., Burdett, T., Fiorelli, B., Fonseca, N.A., Gonzalez-Porta, M., Hastings, E., Huber, W., Jupp, S., Keays, M., Kryvych, N., McMurry, J., Marioni, J.C., Malone, J., Megy, K., Rustici, G., Tang, A.Y., Taubert, J., Williams, E., Mannion, O., Parkinson, H.E. and Brazma, A. Expression Atlas update--a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res* 2014;42(Database issue):D926-932.
- Picard, B., Berri, C., Lefaucheur, L., Molette, C., Sayd, T. and Terlouw, C. Skeletal muscle proteomics in livestock production. *Briefings in functional genomics* 2010;9(3):259-278.
- Picard, B., Lefaucheur, L., Berri, C. and Duclos, M.J. Muscle fibre ontogenesis in farm animal species. *Reprod. Nutr. Dev.* 2002;42(5):415-431.
- Rajesh, R.V., Heo, G.N., Park, M.R., Nam, J.S., Kim, N.K., Yoon, D., Kim, T.H. and Lee, H.J. Proteomic analysis of bovine omental, subcutaneous and intramuscular preadipocytes during in vitro adipogenic differentiation. *Comp Biochem Phys D* 2010;5(3):234-244.
- Rosignol, P., Ho-Tin-Noe, B., Vranckx, R., Bouton, M.C., Meilhac, O., Lijnen, H.R., Guillin, M.C., Michel, J.B. and Angles-Cano, E. Protease nexin-1 inhibits plasminogen activation-induced apoptosis of adherent cells. *The Journal of biological chemistry* 2004;279(11):10346-10356.
- Saleh, J., Summers, L.K., Cianflone, K., Fielding, B.A., Sniderman, A.D. and Frayn, K.N. Coordinated release of acylation stimulating protein (ASP) and triacylglycerol clearance by human adipose tissue in vivo in the postprandial period. *J. Lipid Res.* 1998;39(4):884-891.

- Spalding, K.L., Arner, E., Westermark, P.O., Bernard, S., Buchholz, B.A., Bergmann, O., Blomqvist, L., Hoffstedt, J., Naslund, E., Britton, T., Concha, H., Hassan, M., Ryden, M., Frisen, J. and Arner, P. Dynamics of fat cell turnover in humans. *Nature* 2008;453(7196):783-787.
- Sudre, K., Leroux, C., Pietu, G., Cassar-Malek, I., Petit, E., Lustrat, A., Auffray, C., Picard, B., Martin, P. and Hocquette, J.F. Transcriptome analysis of two bovine muscles during ontogenesis. *J. Biochem. (Tokyo)* 2003;133(6):745-756.
- Sugiyama, Y., Suzuki, A., Kishikawa, M., Akutsu, R., Hirose, T., Waye, M.M.Y., Tsui, S.K.W., Yoshida, S. and Ohno, S. Muscle develops a specific form of small heat shock protein complex composed of MKBP/HSPB2 and HSPB3 during myogenic differentiation. *J. Biol. Chem.* 2000;275(2):1095-1104.
- Sun, S.Y., Zhang, W., Han, X., Huang, R.H. and Shi, F.X. Cell proliferation and apoptosis in the fetal and neonatal ovary of guinea pigs. *Genet. Mol. Res.* 2014;13(1):1570-1578.
- Suzuki, K., Kishioka, Y., Wakamatsu, J. and Nishimura, T. Decorin activates Akt downstream of IGF-IR and promotes myoblast differentiation. *Anim. Sci. J.* 2013;84(9):669-674.
- Taga, H., Chilliard, Y., Meunier, B., Chambon, C., Picard, B., Zingaretti, M.C., Cinti, S. and Bonnet, M. Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown? *Journal of cellular physiology* 2012;227(4):1688-1700.
- Takahashi, Y., Coppola, D., Matsushita, N., Cualing, H.D., Sun, M., Sato, Y., Liang, C., Jung, J.U., Cheng, J.Q., Mule, J.J., Pledger, W.J. and Wang, H.G. Bif-1 interacts with Beclin 1 through UVRAG and regulates autophagy and tumorigenesis. *Nat. Cell Biol.* 2007;9(10):1142-1151.
- Tang, D., Kang, R., Livesey, K.M., Cheh, C.W., Farkas, A., Loughran, P., Hoppe, G., Bianchi, M.E., Tracey, K.J., Zeh, H.J., 3rd and Lotze, M.T. Endogenous HMGB1 regulates autophagy. *J. Cell Biol.* 2010;190(5):881-892.
- Tavallaie, S., Rahsepar, A., Abdi, H., Mouhebaty, M., Moudi, F., Esmaily, H.O., Majid, G.M. and Ferns, G. Association between indices of body mass and antibody titers to heat-shock protein-27 in healthy subjects. *Clin. Biochem.* 2011;44(13):S4-S4.
- Tzen, C.Y. and Huang, Y.W. Cloning of murine early quiescence-1 gene: the murine counterpart of dermatopontin gene can induce and be induced by cell quiescence. *Exp. Cell Res.* 2004;294(1):30-38.
- UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* 2015;43(Database issue):D204-212.
- vanderLoop, F.T.L., Schaart, G., Timmer, E.D.J., Ramaekers, F.C.S. and vanEys, G.J.J.M. Smoothelin, a novel cytoskeletal protein specific for smooth muscle cells. *J. Cell Biol.* 1996;134(2):401-411.
- Welsh, G.I., Griffiths, M.R., Webster, K.J., Page, M.J. and Tavare, J.M. Proteome analysis of adipogenesis. *Proteomics* 2004;4(4):1042-1051.
- Woldt, E., Matz, R.L., Terrand, J., Mlih, M., Gracia, C., Foppolo, S., Martin, S., Bruban, V., Ji, J., Velot, E., Herz, J. and Boucher, P. Differential Signaling by Adaptor Molecules LRP1 and ShcA Regulates Adipogenesis by the Insulin-like Growth Factor-1 Receptor. *J. Biol. Chem.* 2011;286(19):16775-16782.
- Wolfing, B., Buechler, C., Weigert, J., Neumeier, M., Aslanidis, C., Schoelmerich, J. and Schaffler, A. Effects of the new C1q/TNF-related protein (CTRP-3) "cartonectin" on the adipocytic secretion of adipokines. *Obesity (Silver Spring)* 2008;16(7):1481-1486.
- Yamauchi, T., Kamon, J., Minokoshi, Y., Ito, Y., Waki, H., Uchida, S., Yamashita, S., Noda, M., Kita, S., Ueki, K., Eto, K., Akanuma, Y., Froguel, P., Foufelle, F., Ferre, P., Carling, D., Kimura, S., Nagai, R., Kahn, B.B. and Kadowaki, T. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nat. Med.* 2002;8(11):1288-1295.
- Yokota, T., Meka, C.S., Medina, K.L., Igarashi, H., Comp, P.C., Takahashi, M., Nishida, M., Oritani, K., Miyagawa, J., Funahashi, T., Tomiyama, Y., Matsuzawa, Y. and Kincade, P.W. Paracrine regulation of fat cell formation in bone marrow cultures via adiponectin and prostaglandins. *J. Clin. Invest.* 2002;109(10):1303-1310.

Zhang, C.B., He, Y.K., Okutsu, M., Ong, L.C., Jin, Y., Zheng, L., Chow, P., Yu, S., Zhang, M. and Yan, Z. Autophagy is involved in adipogenic differentiation by repressing proteasome-dependent PPAR gamma 2 degradation. *Am J Physiol-Endoc M* 2013;305(4):E530-E539.

Conclusion

Cette étude a permis de générer un protéome de 752 and 514 protéines respectivement pour le tissu adipeux et le muscle, dont 338 communes aux deux tissus. L'analyse des termes GO a révélé les fonctions biologiques correspondant à la prolifération et la différenciation cellulaires dans les deux tissus. Indépendamment, le métabolisme lipidique dans le TA et l'organisation structurale et contractile dans le muscle, ont été identifiés en cohérence avec les résultats des principaux travaux chez le bovin ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)).

Une originalité des résultats est d'identifier l'apoptose parmi les GO les plus représentés pour les deux tissus, en particulier la régulation négative de l'apoptose. Ceci est en accord avec les données de Chaze et al. ([Chaze et al., 2008](#)) indiquant que le nombre total de fibres musculaires est dépendant de l'équilibre entre prolifération cellulaire et apoptose durant les deux premiers trimestres de gestation. Le même phénomène semble être mis en place pour le contrôle du nombre d'adipocytes. La voie Wnt est apparue comme impliquée dans la régulation de la croissance des deux tissus avec respectivement 16 protéines annotées dans le muscle et 21 dans le tissu adipeux.

L'analyse des interactomes met en évidence des protéines telles que HSPB1, GRB2 ou VCAM1 ayant un rôle central dans le développement de chacun des deux tissus.

Les résultats majeurs de l'étude sont apportés par l'identification des protéines potentiellement sécrétées par chacun des deux tissus et pouvant agir sur l'autre tissu. Ainsi, les deux modules développés dans ProteINSIDE, à savoir l'identification de protéines potentiellement sécrétées et l'analyse des interactions entre protéines, permettent de proposer des hypothèses nouvelles quant à la régulation de l'équilibre muscle/tissu adipeux. En particulier le rôle de l'Adiponectine, ou de protéines de la matrice extracellulaire telles que la Décorine et leur potentiel lien avec le phénomène d'autophagie révélé dans nos études, sont à approfondir. Afin, de valider les hypothèses avancées dans cette étude, les protéines candidates pour le contrôle de la régulation du rapport muscle/TA seront analysées dans les deux tissus à l'aide d'anticorps spécifiques à différents stades de la vie foetale (échantillons des thèses de T. Chaze et H. Taga).

La suite de ce travail est de regarder si les protéines centrales ou impliquées dans les processus de développement des tissus ont une abondance modifiée par la nutrition maternelle selon le protocole décrit dans la communication 4. Ce type de résultat fournirait des pistes pour une régulation par la nutrition de l'équilibre de la masse de muscle *versus* masse de tissu adipeux.

**DISCUSSION GENERALE, CONCLUSION ET
PERSPECTIVES**

Discussion générale, conclusions et perspectives

Les travaux réalisés au cours de cette thèse ont conduit à deux avancées majeures. La première est le développement du premier serveur web de fouille de données de génomique dédié aux ruminants. La deuxième est l'apport de connaissances nouvelles sur les protéines impliquées dans les mécanismes de prolifération et de différenciation cellulaires relatifs au muscle et/ou au TA, au cours de la croissance fœtale bovine. Nous avons en outre identifié des protéines et mécanismes moléculaires ou cellulaires, potentiellement clé dans l'ontogenèse des tissus adipeux et musculaires.

1 – L'analyse intégrative de données de génomique chez les ruminants grâce à ProteINSIDE

Le développement et l'assemblage d'outils bioinformatiques ont conduit à la mise en place du serveur web ProteINSIDE. C'est actuellement le seul serveur proposant, **en une seule requête**, une fouille de données de génomique combinant la **synthèse des connaissances biologiques disponibles**, une **annotation fonctionnelle avec des termes GO**, une **prédiction de protéines sécrétées** et une **recherche d'interactions entre protéines** et ceci pour une liste de gènes ou de protéines. Ce serveur est particulièrement utile pour l'analyse des données de génomique chez le ruminant et est complémentaire aux outils existants chez l'homme et les espèces modèles.

1.1 – Utilisation et valorisation de ProteINSIDE

La conception de ProteINSIDE a nécessité environ 7 mois et l'outil est apparu publiquement en octobre 2013, les utilisations précédant cette date étant exclusivement conduites par l'équipe Amuvi. A ce jour, ProteINSIDE a enregistré une croissance dans le nombre d'analyses effectuées et différents pics d'utilisations ont été enregistrés suite aux présentations de l'outil dans des congrès ([Figure 32](#)). En effet ProteINSIDE a été diffusé au travers de 5 communications orales (journées des doctorants 2014-2015 de l'unité, présentation à la journée AUDACE 2014 organisée par le conseil régional d'Auvergne, congrès internationaux IWBBIO 2014 et JOBIM 2015), trois posters (pour le congrès RECOMB 2013, la journée de l'école doctorale 2014 et l'action COST Farm Animal Proteomics 2014) et 3 articles scientifiques (congrès IWBBIO 2014, article GCB 2015 et article PlosONE 2015). ProteINSIDE a aussi été référencé sur le site OmicTools (www.omictools.com). Les ID les plus analysés avec ProteINSIDE proviennent de l'homme et du bovin ([Figure 32](#)).

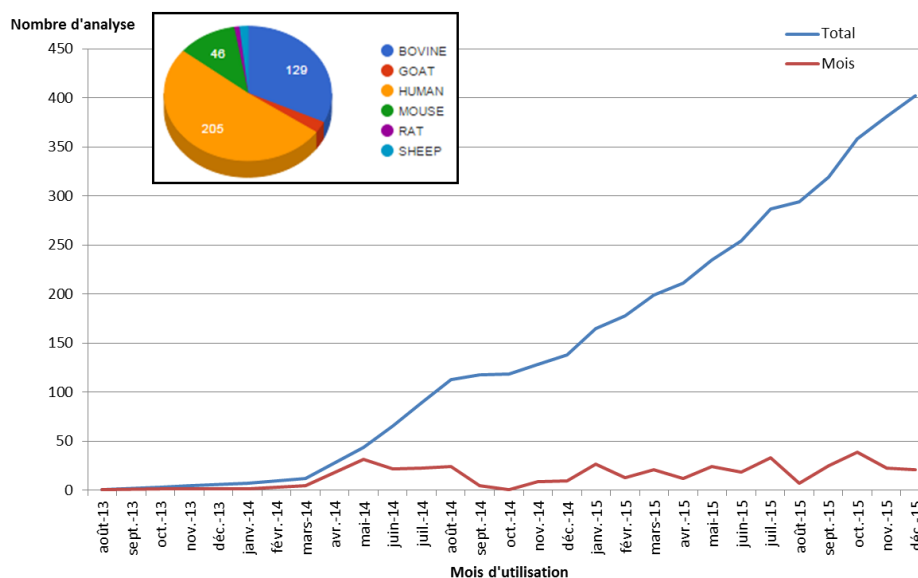


Figure 32: Nombre total des analyses effectuées sur ProteINSIDE au 16 décembre 2015. On peut y observer les différents pics d'activités du site selon les mois. Un diagramme montre la répartition des analyses effectuées pour toutes les espèces.

1.2 – Limites à l'utilisation de ProteINSIDE

À ce jour, des limites subsistent et concernent l'utilisation et l'entretien du serveur web ProteINSIDE.

En ce qui concerne l'utilisation, les analyses créées avec ProteINSIDE sont limitées à 4000 ID. Cette limite a été mise en place afin de limiter le temps d'analyse, car l'outil est capable de gérer plus d'ID. Créer une analyse basique dans l'espèce humaine avec 4000 ID génère un temps de traitement de plusieurs heures (environ 12 heures pour des ID bien annotés). A ce jour, seuls les administrateurs de l'outil peuvent créer et lancer des analyses sans limites d'ID. ProteINSIDE est un serveur qui interroge d'autres outils ou bases de données comme QuickGO, AmiGO, UniProt ou Psiquic. Si l'une de ces ressources venait à être arrêtée, le module associé de ProteINSIDE ne produirait plus de résultats. ProteINSIDE est un outil uniquement disponible en ligne et nécessite donc une connexion à internet pour être utilisé. Enfin, ProteINSIDE a été conçu pour être mis à jour automatiquement tous les mois. Ce processus entretient constamment une base de données d'ID et de termes GO à la dernière version. Cependant, il nécessite une administration mensuelle de la base de données afin de vérifier les éventuelles erreurs. En effet, les bases de données sources (UniProt et NCBI) sont composées de fiches descriptives des protéines ou gènes, possédant souvent des caractères spéciaux parfois non pris en compte dans les programmes de mise à jour. Ces erreurs peuvent stopper le programme de mise à jour et la base de données reste ainsi à la version précédente pour un mois supplémentaire. Ainsi, la pérennité de ProteINSIDE nécessite une administration par une personne physique et ayant des compétences en programmation pour pallier aux différents problèmes engendrés par les mises à jour des outils dont dépend ProteINSIDE.

1.3 – Perspectives d'évolution de ProteINSIDE

Quel que soit l'outil bioinformatique, il existe toujours une multitude d'ajouts possibles pour faciliter l'utilisation et améliorer les performances d'utilisation. La visualisation des résultats, actuellement effectuée par des tableaux dynamiques issus de Google Charts, pourrait être remplacée. En effet, ces tableaux sont parfaitement fonctionnels, mais leur pérennité n'est pas assurée, car ils dépendent de la version actuelle du projet d'affichage dynamique de Google (tableaux et graphiques). Leur utilisation nécessite une connexion constante aux serveurs de Google, or les connexions à Google sont parfois bloquées dans certains pays : un utilisateur chinois a dû contourner le système pour utiliser ProteINSIDE. En outre, la version actuelle du projet d'affichage dynamique de Google subit des mises à jour nécessitant une maintenance de ProteINSIDE. Un système stable comme DataTable (<https://www.datatables.net/>) possède des fonctionnalités similaires aux tableaux Google et ne nécessite pas de connexion constante, car il peut être téléchargé localement.

En 2015, la version Cytoscape web (version programmée en flash) utilisée pour l'affichage des réseaux de ProteINSIDE a été arrêtée. En effet, Cytoscape web utilise désormais une nouvelle version codée en Javascript pour un support complet de tous les navigateurs sans perte de performances (même pour les gros réseaux ; <http://js.cytoscape.org/>). Cette nouvelle version nécessite une reprogrammation complète de l'affichage des réseaux de ProteINSIDE et n'a pas pu être réalisée. Enfin, l'interface web de ProteINSIDE, qui a été entièrement développée manuellement, pourrait être refaite en utilisant des supports plus sécurisés et plus fiables tels que le Framework Symfony (www.symfony.com). Ceci faciliterait la maintenance de l'outil notamment par des utilisateurs autres que les développeurs de l'outil. Ainsi, la gestion des utilisateurs, des résultats et de la base de données, pourrait ainsi être effectuée par du personnel non programmeur.

Chaque module du workflow de ProteINSIDE peut être encore amélioré. Le module « ID Mapping » souffre parfois d'un manque d'identification d'ID envoyé dans l'outil. En effet, de nombreux noms de gènes sont dupliqués à cause de différentes isoformes de protéines, noms de gènes trop similaires entre espèces, ou simple doublon dans la base de données source avec un lettrage différent. Par exemple, pour les ID de la tubuline (TUBB) chez l'homme, les bases de données référencent les ID Q9BUU9, Q6P602, Q96B85 et P07437 repris et proposés par ProteINSIDE. Seul le dernier ID correspond à la tubuline, il est le plus couramment utilisé et le mieux renseigné avec des annotations vérifiées sur la base de données UniProt. Ce problème d'ID est bien connu des bases de données, la base UniProt cherche actuellement à homogénéiser ses fiches d'informations, mais n'apporte pas de solution au problème de duplication. Une solution pour ProteINSIDE serait de demander aux utilisateurs d'envoyer des séquences au lieu des ID pour une nouvelle analyse. Un Blast sur les bases de données du NCBI (UniProtKB/Swiss-Prot par exemple) déterminerait alors l'ID unique à la séquence. Le module de Gene Ontology pourrait proposer une représentation graphique hiérarchisée des termes GO en diagrammes dégressifs comme le propose l'outil Krona ([Ondov et al., 2011](#)). Cette visualisation est plus allégée qu'un grand tableau et moins grande que les réseaux proposés actuellement par ProteINSIDE. Cependant, la limite de l'outil Krona est qu'il réalise des diagrammes monoparentés alors que la hiérarchisation des termes GO nécessite des diagrammes multiparentés. Ainsi, une adaptation de Krona, ou l'utilisation d'un autre outil, améliorerait la visualisation des relations entre termes GO. La prédiction des protéines sécrétées, par des voies de sécrétion qui n'impliquent pas de peptide signal (voies non classiques de sécrétion), pourrait être

incrémentée avec l'outil SecretomeP ([Bendtsen et al., 2004](#)). Même si cet outil produit les mêmes résultats que SignalP (prédiction des protéines sécrétées par un peptide signal), il propose aussi un score qui indique si la protéine est sécrétée par une voie non classique. Avec cette implémentation, ProteINSIDE serait le premier serveur combinant autant de méthodes de prédiction des protéines sécrétées : les termes GO et les prédictions de peptides signaux ou d'autres voies de sécrétion, et de la localisation cellulaire de la protéine. Pour finir, une amélioration importante de ProteINSIDE serait la possibilité de comparer les résultats de deux analyses. Pour ce qui concerne la croissance du muscle relativement à celle du TA, les ID correspondant aux protéomes musculaires et adipeux pourraient être comparés en temps réel. Les termes GO similaires et spécifiques à chaque tissu pourraient être extraits, ainsi que les protéines annotées. De la même façon les protéines sécrétées pourraient être triées par tissu. Enfin, superposer des réseaux d'interactions des deux tissus serait un plus pour mettre en évidence les fonctions similaires entre chaque tissu lors de la croissance fœtale. Dans un contexte plus général, ProteINSIDE pourrait ouvrir son potentiel d'analyse à plusieurs autres espèces d'intérêt agronomique comme le poulet, le porc et le lapin. ProteINSIDE pourrait aussi être décliné en versions dédiées aux espèces végétales (comme le maïs ou le blé) et bactériennes pour lesquelles nous avons eu des demandes.

2 – Les protéomes musculaire et adipeux au cours de la croissance fœtale bovine

Notre objectif était de mettre en évidence les processus moléculaires et biologiques accompagnant la croissance par hyperplasie et par hypertrophie des tissus musculaires et/ou adipeux, et de révéler des interactions pour la régulation de cette croissance. Nous avons utilisé les protéomes établis par les études de Thibault Chaze et Hajer Taga sur la croissance fœtale des tissus musculaire et adipeux (respectivement composés de 120 et 143 protéines bovines ; ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#))). Nous avons complété les protéomes musculaire et adipeux en identifiant respectivement de 514 et 752 protéines bovines présentes dans les tissus fœtaux à 140 jpc. La compilation de ces données constitue un jeu de données unique à ce jour, composé de 556 et 801 protéines, respectivement pour le muscle et le TA. Grâce à ProteINSIDE nous avons pu fouiller ces données pour apporter des connaissances nouvelles. Nous présentons ici les avancées scientifiques et méthodologiques majeures issues de ce travail et nous en discutons les limites.

2.1 – Apports de connaissances par l'analyse du protéome fœtal

La comparaison des 556 et 801 protéines musculaires et adipeuses a révélé que 374 protéines sont communes. La fouille de données relative aux protéines communes puis aux protéines spécifiques à chaque tissu a révélé des fonctions biologiques communes et spécifiques à la croissance des deux tissus. Elle a aussi permis de décrire pour la première fois les protéines sécrétées par les tissus fœtaux.

Les fonctions biologiques nouvellement identifiées dans les protéomes musculaires et adipeux

Les fonctions biologiques majeures associées à la croissance fœtale des tissus adipeux et musculaires sont l'autophagie, une régulation négative de l'apoptose, la voie de signalisation des protéines Wnt et la prolifération cellulaire.

Des protéines de l'autophagie ont été identifiées comme centrales dans les réseaux PPI réalisés entre les protéines fœtales et les protéines déclarées dans les bases de données de PPI (réseaux étendus), et ceci à la fois pour les protéomes caractérisant la croissance fœtale ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Kaspric et al., 2015](#); [Taga et al., 2012](#)), l'âge 140 jpc (article 5 en préparation) et le protéome total. Des interactions directes entre les protéines fœtales et GABARAPL1, GABARAPL2, GABARAP, MAP1LC3A et MAP1LC3B ont été établies dans un premier temps ([Kaspric et al., 2015](#)) puis complétées par l'identification de 70 protéines impliquées dans des processus autophagiques dans le PPI étendu du protéome total (Article 5 en préparation). Parmi elles, 58 protéines sont impliquées dans la macroautophagie. Ce réseau étendu comporte les 46 protéines centrales dont GABARAPL2, GABARAPL1 et MAP1LC3A. Selon la base de données Expression Atlas ([Kapushesky et al., 2010](#)), GABARAPL1 et MAP1LC3B sont connues pour être synthétisées par le TA et GABARAPL1, GABARAPL2 et MAP1LC3B sont synthétisées par le muscle. Les protéines fœtales qui ont de nombreuses interactions avec les protéines de l'autophagie sont notamment HSPB1, FKBP4, TCP1, STP1. L'hypothèse d'un rôle important de l'autophagie pour la croissance fœtale des tissus musculaires et adipeux est soutenue par des résultats obtenus essentiellement sur des modèles murins. Des modifications dans l'expression des gènes associés à l'autophagie ont lieu au cours de la différenciation de la lignée musculaire C2C12 et adipeuse 3T3-L1. En particulier, l'abondance d'Atg7 augmente pendant la différenciation, et l'inhibition de l'expression du gène Atg7 empêche la fusion des myoblastes et leur différenciation ([McMillan et Quadriatero, 2014](#)). Ces auteurs suggèrent que l'autophagie protègerait les myoblastes en différenciation de l'apoptose. Dans les 3T3-L1, l'induction de l'autophagie serait nécessaire à la prolifération et à la différenciation des préadipocytes en stimulant le remodelage cellulaire et en diminuant le nombre de mitochondries ([Skop et al., 2014](#)) indispensables au processus d'adipogenèse ([Gregoire et al., 1998](#)). Ce rôle de l'autophagie dans le remodelage cellulaire des préadipocytes pourrait expliquer la forte liaison de TCP1 que nous avons identifié dans le protéome fœtal adipeux, avec les protéines de l'autophagie. En effet, TCP1 est un médiateur du repliement des protéines dans le cytoplasme, en particulier de l'actine et de la tubuline. L'abondance de TCP1 est fortement associée à la prolifération d'une grande variété de cellules humaines et murines (fibroblastes, macrophages, carcinome, lymphome, myélome ; ([Yokota et al., 1999](#))). Enfin, le rôle central de l'autophagie dans l'ontogenèse des TA a été prouvé par la délétion du gène Atg5 chez la souris ([Baerga et al., 2009](#)). Les nouveau-nés issus de souris Atg7^{-/-}, ont seulement 15% d'adipocytes différenciés dans le TA sous-cutané comparativement à des souris sauvages.

L'analyse du protéome total (regroupant les protéines des deux études), a permis de révéler les principales fonctions biologiques communes aux tissus musculaires et adipeux comme la prolifération cellulaire, une régulation négative de l'apoptose, et la voie de signalisation Wnt ([Figure 33](#)).

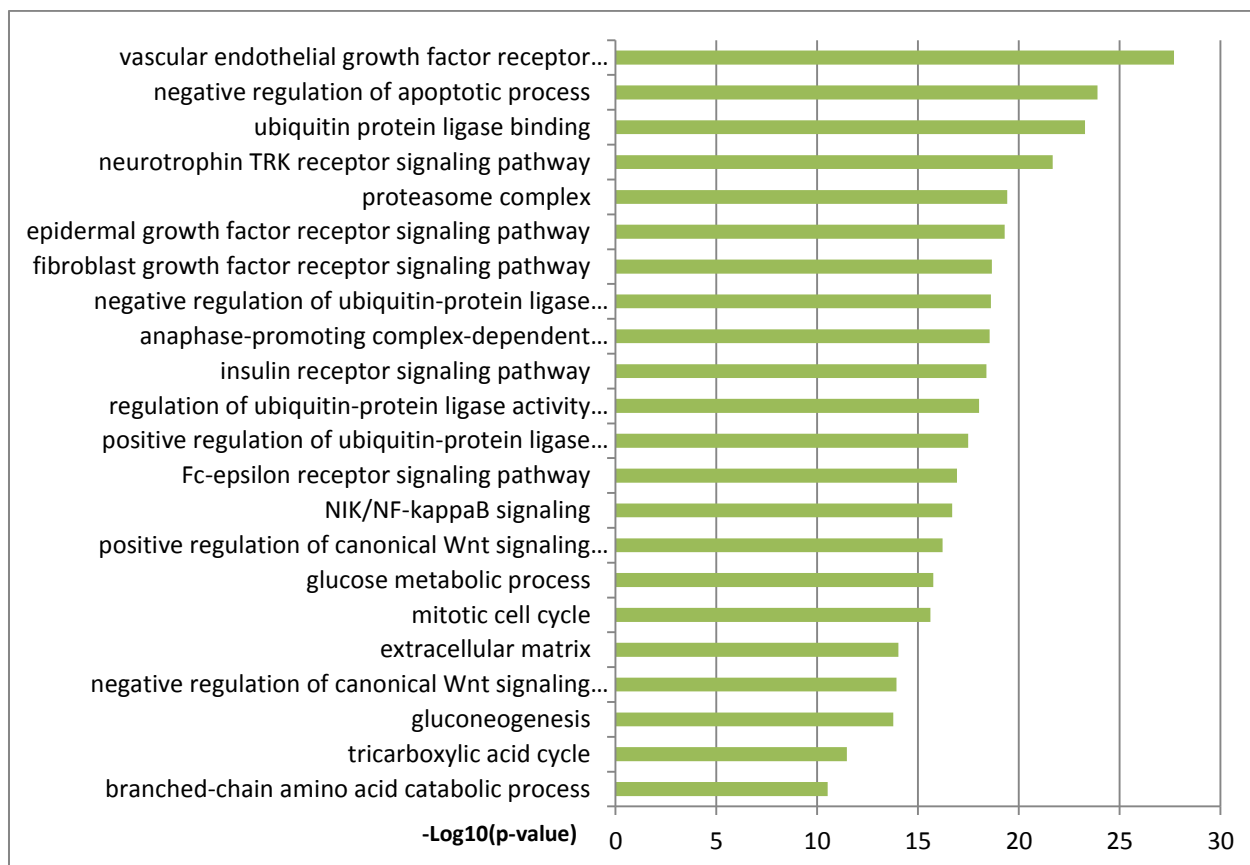


Figure 33 : Les fonctions biologiques les plus enrichies dans le protéome total et commun aux tissus adipeux et musculaires fœtaux.

Pour la première fois chez le bovin (Figure 33), nous avons identifié de nombreuses protéines annotées par des termes GO relatifs à la prolifération cellulaire (PA2G4, PDXK, SERPINF1, KRT16, PDAP1, TXN et FSCN1) et sa régulation positive (KRT6A, NUDT16, DNAJA2, RPRD1B et CRKL) ou négative (MAP2K1, CTBP1, WARS, COPS8 et NCK2). Toutes les protéines précédentes ont déjà été identifiées lors d'expérimentation *in vitro* sur des lignées musculaires C2C12 et adipeuses 3T3L1 (cf. Etude Bibliographique, paragraphe 4, tableau 3). Outre ces protéines, nous avons identifié pour la première fois *in vivo* 13 et 36 protéines spécifiques respectivement des tissus musculaires et adipeux, et susceptibles de participer à la prolifération des progéniteurs cellulaires. L'analyse de ces données indique aussi que nous avons identifié un plus grand nombre de protéines impliquées dans la régulation de la prolifération des progéniteurs adipeux que dans celle des myoblastes. Pour le muscle, nous avons identifié un plus grand nombre de protéines impliquées dans la différenciation des myotubes. Ceci est en cohérence avec la chronologie de croissance des tissus musculaire et adipeux décrite chez le bovin (Bonnet *et al.*, 2015).

Nos données ont aussi identifié de nombreuses protéines annotées par des termes GO relatifs à l'apoptose (59 protéines pour le muscle et 71 pour le TA) et en particulier à une régulation négative de l'apoptose (34 protéines annotées pour le muscle et 43 pour le TA). Ces résultats sont en accord avec ceux de Chaze *et al.* (Chaze *et al.*, 2008) qui avaient montré une augmentation de l'abondance de protéines ayant une fonction anti apoptotique dans le muscle ST entre 110 et 180 jpc, comme par exemple Hsp27. Ils en avaient conclu que la détermination du nombre total de fibres

musculaires devait être dépendante d'un équilibre entre prolifération cellulaire et apoptose durant les deux premiers trimestres de gestation chez le bovin. Cette hypothèse est cohérente avec l'activité anti-apoptotique de la kinase Akt induite durant la myogenèse qui permettrait la survie de myoblastes C2C12 différenciés ([Kamradt et al., 2002](#)). Toutefois, les voies anti-apoptotiques impliquées dans le muscle durant la myogenèse restent peu connues.

Enfin, nous avons identifié des protéines adipeuses et musculaires impliquées dans la voie de signalisation Wnt. Ces résultats sont cohérents avec les fonctions connues des glycoprotéines Wnt. En effet, l'activation de la voie de signalisation Wnt induit la myogenèse et inhibe l'adipogenèse ([Du et al., 2010](#)). L'activation de la voie de signalisation Wnt stabilise la b-caténine. Or, Wnt/ b-caténine module l'expression des facteurs de transcription myogéniques et adipogéniques clés, Pax 3, Myf5 et PPAR γ . La surexpression de la b-caténine augmente la prolifération de progéniteurs musculaires ([Otto et al., 2008](#)) et inhibe la différenciation des préadipocytes ([Liu et Farmer, 2004](#)). Les protéines fœtales, que nous avons identifiées et qui sont annotées par des termes GO relatifs à la voie de signalisation Wnt, sont principalement les protéines du complexe du protéasome et elles sont numériquement plus abondantes dans le TA fœtal que dans le muscle (PSMA1, PSMA6, PSMB2, PSMB3, PSMB4, PSMB6, PSMB7, PSMC1, PSMC2, PSMD9, PSME1 et PSME2). Ce résultat est probablement à relier au rôle important du protéasome dans la régulation de l'adipogenèse. En effet, dans le modèle murin, le protéasome dégrade les facteurs de transcription adipogéniques ubiquitinés, PPAR γ et CEBP, et favorise ainsi le maintien de la prolifération au détriment de la différenciation adipocytaire (2 ref cf la publi). Nous avons aussi identifié la protéine TBL1XR1 (F-box-like/WD repeat-containing protein) comme impliquée dans la voie de signalisation Wnt, mais dont le rôle dans le TA et le muscle reste à identifier.

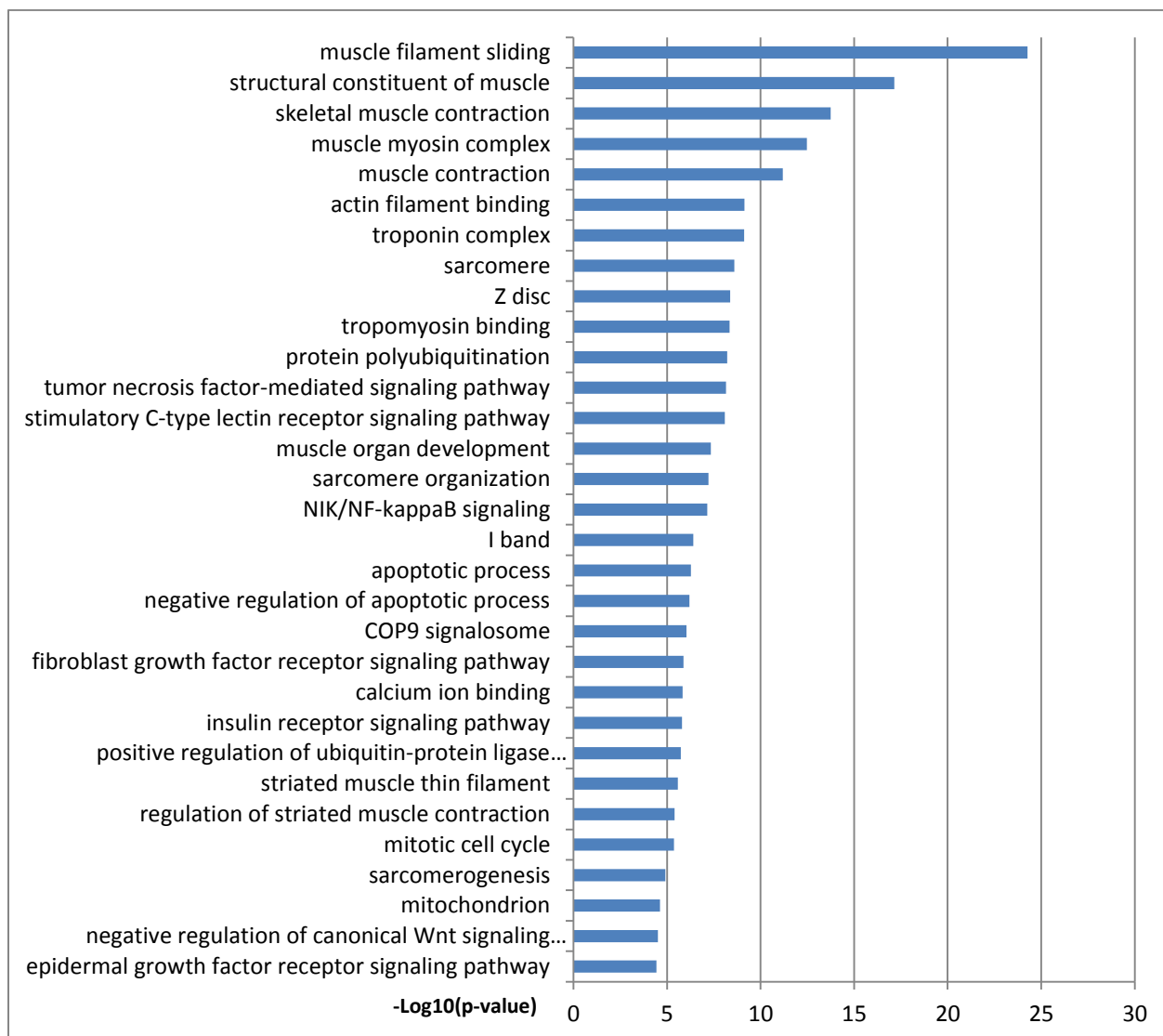


Figure 34 : Les fonctions biologiques les plus enrichies dans le protéome spécifique au tissu musculaire fœtal.

Enfin, nous avons identifié des fonctions biologiques qui sont spécifiques des tissus musculaires et adipeux (Figure 34 et Figure 35). Pour le muscle, les 10 fonctions biologiques les plus enrichies correspondent à l'organisation ultrastructurale des fibres musculaires et à la mise en place de l'appareil contractile du muscle. Ceci est cohérent avec le fait que la différenciation contractile et métaboliques des fibres primaires s'achève entre 110 et 180 jours de gestation, et avec la synthèse des protéines myofibrillaires décrite dans des études précédentes (Chaze *et al.*, 2008; Picard *et al.*, 2002). En outre, nous avons identifié des termes GO relatifs à la prolifération cellulaire (SRA1, GOLPH3, CDV3, CCKBR, BIN1 et WDR12), notamment sa régulation positive (MVD, MYC et CCKBR) et négative (FABP3). Nous avons identifié quelques protéines impliquées dans la différenciation cellulaire (SRA1) et sa régulation positive (PPP2CA). De plus, quatre protéines sont connues pour être spécifiquement impliquées dans la différenciation des cellules musculaires (MAP2K6 et SYNE1), dans la régulation positive (HIF1AN) et dans la régulation négative (ANKRD2). Nous avons identifié dans le muscle fœtal de nombreuses protéines annotées par des termes GO relatifs à l'apoptose (LMNA, PSMD7, PAK2, PSMA7, PSMC6, FADD, SEPT4, DSG1, PPP2CA et UBA52) et notamment dans sa

régulation négative (PSMC6, PAK2, BAG3, MYC, PSMD7, PSMA7, GOLPH3, LTF et UBA52). Enfin, nous avons identifié des protéines impliquées dans la voie de signalisation Wnt (MYC et PPP2CA) et en particulier dans sa régulation positive et négative (PSMA7, PSMC6, PSMD7 et UBA52).

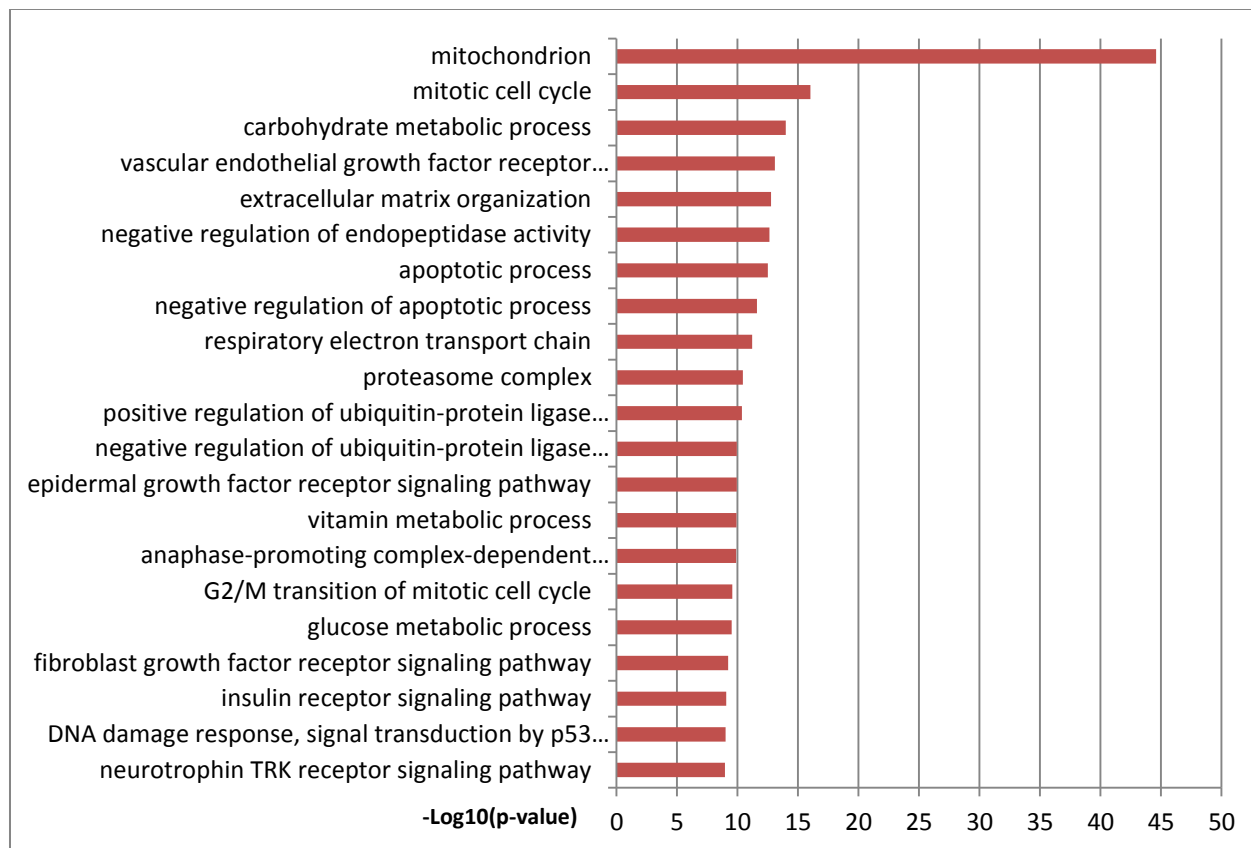


Figure 35 : Les fonctions biologiques les plus enrichies dans le protéome spécifique du tissu adipeux fœtal.

Pour le TA, les fonctions biologiques les plus enrichies correspondent à la mitose et au cycle cellulaire, mais aussi au métabolisme. Ceci semble cohérent avec la forte activité de prolifération des TA en particulier à 140 jpc, et avec le début de différenciation des progéniteurs adipeux. Enfin, les termes GO les plus enrichis correspondent aussi à la mitochondriogenèse et au remodelage de la matrice extracellulaire, deux événements cellulaires clés de l'adipogenèse. Pour ces fonctions biologiques, les protéines spécifiques au TA fœtal ont été listées. Nous avons identifié des protéines impliquées dans la prolifération cellulaire (RBBP7, NUDC, HDGFRP3, SRRT, APPL2 et DCTN2), notamment sa régulation positive (PRDX3, NPM1, CNBP et CAPNS1) et négative (ALDH1A2, CSNK2B, PLG, SOD2, NPM1 et CTBP2). La protéine SERPINF2 est impliquée dans la différenciation cellulaire et la protéine CTBP2 dans la différenciation du TA blanc. Nous avons identifié des protéines impliquées dans le métabolisme des acides gras (HAACL1, HADHB, HADHA, ACADL, PCCB, ACADVL et ACAD8), l'apoptose (ADD1, YARS, LMNB1, PSMD10, PRDX3, KPNB1, YWHAB, PKP1, PSMA3, PSMD5, AIFM1, PSMD8, CASP4, MAPK1, YWHAG, DPF2, PSMB1, PRDX5, BID, PSMA2, PSMD13 et VIM) et plus précisément dans la régulation négative (PSMA2, PRDX5, PSMD8, NPM1, PSMD10, FHL2, FLNA, PSMD5, DDX3X, PRDX3, GSTP1, HSP90B1, CAT, PSMD13, PSMB1, PSMA3, KRT18, FABP1 et CFL1).

Enfin, nous avons identifié des protéines impliquées dans la voie de signalisation Wnt (DDX3X et RAB5A) et en particulier dans sa régulation positive et négative (PSMD10, PSMA2, PSMD13, PSMB1, PSMA3, PSMD5 et PSMD8).

Les protéines sécrétées par les tissus adipeux et musculaires et leur implication dans le dialogue muscle-TA

Pour la première fois, nous produisons une liste de protéines sécrétées par les deux tissus chez des fœtus bovins. Nous avons listé 47 et 93 protéines potentiellement sécrétées par le muscle et le TA, dont 24 protéines communes aux deux tissus. Pour certaines d'entre elles la mise en relation des connaissances sur leur fonction biologique (synthétisée par le module « ID mapping » de ProteINSIDE) et les processus biologiques discutés précédemment suggère leur contribution à la régulation de la croissance des tissus. Par exemple, pour les protéines sécrétées par les deux tissus, la protéine RBP4 (Retinol-binding protein 4) est connue pour jouer un rôle dans le développement embryonnaire ([Quadro et al., 2005](#)), et dans celui du muscle cardiaque ([Wendler et al., 2003](#)). La protéine OGN (Mimecan), est connue pour réguler négativement la prolifération du muscle ([Arita et al., 2002](#)). La protéine APOH régulerait négativement l'apoptose musculaire ([Lin et al., 2005](#)). La protéine KNG1 (Kininogen-1) est impliquée dans la régulation positive de l'apoptose ([Al-Fakhri et al., 2003](#)). Nous avons aussi identifié des protéines sécrétées par le muscle comme la protéine LGALS1 (galectine-1), impliquée dans la régulation l'apoptose ([Bruneel et al., 2005](#)) et dans la différenciation des myoblastes (annotation électronique selon UniProt), et aussi la protéine C1QTNF3 (Complément C1q nécrose tumorale facteur lié à la protéine 3), connue pour être impliquée dans la différenciation des cellules adipeuses ([Kopp et al., 2010](#)) et dans la régulation positive de la sécrétion de l'adiponectine ([Wolfing et al., 2008](#)). Parmi les protéines sécrétées par le TA, nous avons identifié la protéine DCN, connue pour être impliquée dans le développement du muscle squelettique (en particulier dans le muscle squelettique bovin; ([Nishimura et al., 2002](#))), et qui régulerait positivement l'autophagie ([Buraschi et al., 2013](#)). La protéine SERPINF2 (Alpha-2-antiplasmine) participe à la régulation positive de la prolifération des cellules musculaires et régulerait positivement la différenciation cellulaire ([Kanno et al., 2008](#)).

Outre leur originalité, ces données nous ouvrent l'opportunité de proposer des interactions entre le TA et le muscle au cours de la croissance fœtale. Pour cela, nous avons développé une stratégie originale qui consiste à rechercher des interactions physiques (PPI) et des interactions de signalisation (réseaux de signalisation proposés par différentes ressources comme Pathway Studio) entre les protéines sécrétées par un tissu et les protéines membranaires ou présentes dans l'autre tissu. Ainsi, nous avons établi une liste de 44 et 82 protéines de surface, respectivement pour le muscle et le TA. Les analyses des listes de protéines sécrétées et des listes de protéines musculaires ont identifié des dizaines d'interactions, parmi elles nous avons focalisé sur 4 voies de signalisation.

Pour les protéines adipeuses susceptibles de modifier la croissance musculaire, nous avons identifié la voie de signalisation impliquant la protéine adipeuse CLEC3B (Tetranectin) connue pour réguler positivement ENO1 (Alpha-enolase) ([Diaz-Ramos et al., 2012](#)) dans le muscle. Cette voie de signalisation implique un intermédiaire plasmatique, la protéine PLG (synthétisée par le foie et identifiée parmi les protéines musculaires). Or, ENO1 a été décrite comme présente dans le développement précoce du tissu musculaire bovin ([Chaze et al., 2008](#)). Elle est également connue

pour être impliquée dans la régulation négative de la croissance cellulaire ([Ghosh et al., 1999](#)). La deuxième voie de signalisation implique la protéine DCN (Decorin), sécrétée par le TA et susceptible de réguler la protéine THBS1 (Thrombospondin-1), identifiée dans le muscle. La THBS1 est une glycoprotéine impliquée dans des interactions entre les cellules et la matrice extracellulaire pour moduler, de façon cellule-spécifique, la migration et la prolifération cellulaire. L'expression du gène THBS1 a été rapportée dans le tube neural et les myoblastes en différenciation ([Iruelaarispe et al., 1993](#)). Nous avons retrouvé cette protéine dans la liste des protéines à activité anti-apoptotique dans le muscle et dans le TA, ce qui est en accord avec les résultats de ([Pallero et al., 2008](#)) montrant une implication de la protéine THBS1 dans la régulation négative des processus apoptotiques. Ces auteurs, sur des fibroblastes en culture, ont montré que THBS1 active Akt et diminue la voie apoptotique induite par la caspase 3. Toutefois, son implication dans la myogenèse ou l'adipogenèse n'a pas été décrite. Sur des cellules C2C12, Kishioka et al. ([Kishioka et al., 2008](#)) ont montré que DCN stimule la voie TGF-beta par le biais du récepteur à la lipoprotéine (LRP-1) dans des myoblastes non différenciés. Par contre, dans des myoblastes différenciés, DCN inhibe la voie TGF-beta. Ces données montrent que le rôle de protéines de la matrice extracellulaire telles que DCN peut être modifié au cours de la myogenèse. Il serait donc intéressant de suivre l'abondance de cette protéine à différents stades de la myogenèse et adipogenèse bovines afin de mieux comprendre son implication. D'autres auteurs ont suggéré que DCN agirait comme molécule de signalisation à travers une activation de Akt ([Suzuki et al., 2013](#)). Enfin, la protéine DCN est aussi connue pour être impliquée dans la régulation positive de l'autophagie de cellules endothéliales ([Buraschi et al., 2013](#)), rôle à vérifier dans le muscle et le TA.

Pour les protéines musculaires susceptibles d'impacter la croissance adipeuse, nous avons identifié la voie de signalisation impliquant la protéine C1QTNF3 (Complement C1q tumor necrosis factor-related protein 3), identifiée dans le muscle et sécrétée par le muscle, avec ADIPOQ (Adiponectin), synthétisée et sécrétée par le TA. La protéine C1QTNF3 est connue pour être impliquée dans la régulation positive de la sécrétion de l'adiponectine et la régulation de la différenciation des adipocytes. Or, ADIPOQ est connue pour être impliquée dans la régulation négative de la différenciation des cellules adipeuses ([Yokota et al., 2002](#)) et de la prolifération des cellules musculaires ([Arita et al., 2002](#)). ADIPOQ est aussi une adipokine qui agit sur le TA et d'autres tissus pour une régulation du métabolisme des lipides et de la sensibilité à l'insuline ([Yamauchi et al., 2002](#)). La deuxième voie de signalisation implique la protéine DPT (Dermatopontin), présente et sécrétée par le muscle, qui inhibe la protéine DCN, identifiée dans le TA. La protéine DPT est connue pour être impliquée dans la régulation négative de la prolifération cellulaire ([Tzen et Huang, 2004](#)). Cette première analyse *in silico* des interactions entre tissus est à compléter et à affiner. Il en ressort néanmoins que ces interactions impliquent des protéines de la matrice extracellulaire, importantes pour le remodelage cellulaire qui accompagne la myogenèse et l'adipogenèse. Les protéines de la matrice extracellulaire sont aussi importantes dans le dialogue cellule-cellule et dans la signalisation moléculaire de nombreuses voies ([Arnold et Winter, 1998](#); [Gregoire et al., 1998](#)).

2.2 – Les limites de notre étude

Les limites concernent la partie expérimentale d'identification du protéome et les analyses informatiques des données.

Les limites de la technique de protéomique sont le nombre et le type de protéines identifiées. La compilation de 3 expérimentations de protéomique ([Chaze et al., 2008](#); [Chaze et al., 2009](#); [Taga et al., 2012](#)) a généré un jeu de données de 556 et 801 protéines, respectivement pour le muscle et le TA. Or, les techniques actuelles de protéomique telles que le shot-gun permettent d'identifier plus d'un millier de protéines en une seule analyse ([Brambilla et al., 2013](#); [Le Bihan et al., 2012](#)). Le faible nombre de protéines que nous avons identifiées résulte essentiellement de l'électrophorèse 2D. Cette méthode de protéomique ne sépare qu'un nombre restreint de protéines : les plus abondantes (selon l'analyse informatique du gel) et celles d'un poids moléculaire compris entre 10 à 250 kDa. De plus, seules les protéines avec un point isoélectrique compris entre 4 et 8 ont été séparées sur les gels. Ainsi, les protéines qui ont un poids moléculaire trop important et/ou un pI plutôt basique (> 7 unités de pH) ne sont pas séparées et donc n'ont pas été identifiées. Par exemple, les protéines de l'autophagie que nous proposons comme importantes dans l'ontogenèse des tissus n'ont pas été détectées par nos expérimentations car elles ont un pH situé autour de 8 et un poids moléculaire d'environ 14 kDa. Par ailleurs, dans nos études, seules les protéines identifiées avec un bon score Mascot et par au moins trois peptides ont été retenues, ce qui nous a conduits à exclure un grand nombre de protéines, auxquelles s'ajoutent les protéines membranaires mal extraites par nos conditions d'extraction.

En ce qui concerne les analyses bioinformatiques plusieurs limites sont à considérer. Nous avons identifié des protéines bovines, que nous avons analysées avec leurs homologues humains en utilisant les liens d'homologies avec l'espèce humaine. Dans cette démarche, le risque est que les résultats prédictifs obtenus aient une signification biologique limitée dans l'espèce bovine. Cependant, cette démarche est justifiée par le fait que le nombre d'annotations dans l'espèce bovine est très faible par rapport à l'homme. Par exemple, pour le protéome commun bovin (composé de 349 protéines) 183 protéines sont annotées par 619 termes GO (sans utilisation d'annotations électroniques), alors que parmi leurs orthologues humains (331 protéines) 324 protéines sont annotées par 1921 termes GO. Bien que des liens d'homologies entre les protéines soient référencés sur la base HomoloGene, rien ne prouve que les connaissances pour une protéine humaine soient transposables à ses homologues. Une réponse à cela serait la comparaison de séquences protéiques ([Pearson, 2013](#)). Des séquences homologues ont des structures similaires et souvent des fonctions similaires (les génomes bovin et humain auraient environ 95% de gènes en communs ; [Zimin et al., 2009](#)). C'est la raison pour laquelle, ProteINSIDE propose à l'utilisateur de regarder le pourcentage d'homologie des séquences protéiques entre les espèces. De plus, le concept de « fonction biologique » est ambigu car la même enzyme peut avoir des rôles différents dans deux tissus, aussi les annotations pour être pertinentes nécessitent d'être resituées dans le contexte de « la biologie » d'un tissu. Ces constats mettent en évidence que ces analyses informatiques sont une aide précieuse pour prédire des mécanismes ou proposer des hypothèses, mais que de telles hypothèses doivent être confrontées à l'expertise biologique et nécessitent d'être prouvées expérimentalement.

Une autre limite à notre analyse informatique est la recherche *in silico* d'interactions entre tissus. En effet, la liste des protéines de surface référencées dans la base Cell Surface Protein Atlas est limitée à 1492 protéines humaines et il s'agit de protéines susceptibles d'être sur toutes les membranes cellulaires, y compris les intracellulaires. Aussi, il serait pertinent de comparer cette liste avec les protéines identifiées dans les termes GO:0009986 (« cell surface » ; 950 protéines humaines annotées) et GO:0007166 (« cell surface receptor signaling pathway » ; 4300 protéines humaines annotées), pour refaire une analyse de ces interactions entre tissus. Nous avons utilisé uniquement

les protéines potentiellement sécrétées *via* un peptide signal, pourtant nous avons analysé d'autres protéines potentiellement sécrétées par d'autres voies. Ces analyses sont à enrichir avec les protéines qui sont sécrétées par d'autres voies que celle impliquant les peptides signaux. Cette première analyse des interactions entre tissus et donc à enrichir selon les pistes proposées ici.

3 – Conclusions et application du travail de thèse

À l'issue de ce travail de thèse, nous avons conçu un serveur web original et unique pour la fouille de données de génomique. Nous avons constitué des jeux de données uniques sur les protéomes fœtaux des tissus musculaire et adipeux, respectivement composés de 556 et 801 protéines bovines. Le protéome commun à ces deux tissus est quant à lui composé de 374 protéines bovines.

L'apport de ProteINSIDE à la communauté scientifique est un gain de temps considérable pour la fouille de données de génomique par la mise en relation d'outils et de bases de données biologiques. Grâce à ProteINSIDE, nous avons mis en relation de nombreuses connaissances biologiques afin de décrire ces protéomes. Ainsi, nous proposons que des protéines que nous avons identifiées et qui sont annotées dans les GO liés à l'autophagie, à la régulation négative de l'apoptose, à la prolifération cellulaire et à la voie de signalisation Wnt, sont clés dans la croissance des TA et des muscles. Nous rapportons les premières listes de protéines potentiellement sécrétées chez le bovin en période fœtale. De telles connaissances sont les bases nécessaires et indispensables à la compréhension des interactions entre le TA et le muscle. Nous avons proposé une méthodologie pour la recherche *in silico* d'interactions entre les tissus, utilisant les résultats de prédiction de protéines sécrétées de ProteINSIDE et les protéines de surface référencées dans des bases de données et articles scientifiques. Cette méthodologie est néanmoins à améliorer et pourrait être appliquée à des jeux de données de protéomique plus grand pour avoir une description plus exhaustive des interactions.

Différentes applications de ce travail sont envisagées. La première est d'étudier l'implication de l'autophagie dans l'ontogenèse des tissus. Pour cela, il faut tout d'abord vérifier l'expression de gènes spécifiant les protéines de l'autophagie dans les tissus musculaires et adipeux fœtaux et mettre en regard les variations de l'expression des gènes avec les étapes de la myogenèse et de l'adipogenèse.

Une seconde application du travail de thèse serait de vérifier si les protéines que nous proposons comme importantes pour l'ontogenèse des tissus sont spécifiées par des gènes associés à des QTLs (« Quantitative Trait Locus » ou locus de caractères quantitatifs). Brièvement, un QTL est une région de l'ADN (regroupant souvent plusieurs gènes) associée à un caractère phénotypique (tendreté de la viande par exemple). A ce jour la base de données CattleQTLdb ([Hu et al., 2013](#)) référence 36700 QTLs sélectionnés par une expertise manuelle à partir de 620 publications. La mise en relation de QTLs avec les connaissances sur la fonction et la variabilité d'abondance des protéines permettrait de mieux caractériser sur le plan fonctionnel les régions QTLs responsables de la variabilité d'un caractère quantitatif, afin d'avoir de solides hypothèses quant au gène causal sous-jacent. Aussi, j'ai récemment ajouté un module à ProteINSIDE pour identifier les gènes spécifiant les protéines que nous avons identifiées et qui ont été localisés dans les QTLs. Ce module n'est pour le

moment accessible qu'aux administrateurs de ProteINSIDE. Ce module n'a été utilisé qu'à titre « expérimental » sur les données des protéomes fœtaux. Pour les gènes spécifiant les 556 protéines bovines du muscle, 461 gènes sont associés à des QTLs liés à des caractères aussi variés que la reproduction, la production de lait ou à la santé, par exemple. Les 42 QTLs liés aux qualités des carcasses et des viandes, concernaient 433 gènes qui spécifient les protéines identifiées dans le protéome musculaire. Parmi ces gènes, 16 ont été cartographiés dans des QTLs liés au poids de la carcasse, 26 pour la tendreté de la viande et 6 pour sa flaveur. De la même façon, parmi les gènes spécifiant les 801 protéines bovines du TA, 678 ont été cartographiés dans des QTLs. Les 44 QTLs liés aux qualités des carcasses et des viandes, concernaient 639 gènes qui spécifient les protéines identifiées dans le protéome adipeux. Parmi ces gènes, 211 ont été cartographiés dans des QTLs liés au rendement en viande maigre, 70 liés au TA intramusculaire, 25 au pourcentage de TA intermusculaire et 30 à l'épaisseur du gras de la 12^{ème} côte. L'utilisation de ce nouveau module va permettre une étude plus approfondie des protéines associées à des caractères d'intérêt et permettra d'inclure ces informations robustes dans les critères de choix des biomarqueurs de qualité de la viande.

Enfin, une dernière valorisation de ce travail de thèse serait d'utiliser le serveur web et les méthodes analytiques développées pour analyser les données rendues disponibles grâce à la base de données Fat&MuscleDB ([Tournayre et al., 2015](#)). Cette base synthétise des données de transcriptomique et de protéomique relatives à la croissance par hyperplasie ou par hypertrophie des tissus adipeux et musculaire dans différentes espèces. Aussi, l'utilisation de plusieurs milliers d'identifiants associés à l'un de ces événements cellulaires, permettait une compréhension plus robuste des mécanismes moléculaires impliqués dans la croissance des TA et des muscles et dans le dialogue entre ces deux tissus.

ANNEXES

Annexe A.1 : Documents supplémentaires de l'article de Genomics and Computational Biology

Protein function easily investigated by genomics data mining using the ProteINSIDE online tool

Nicolas KASPRIC^{1, 2}, Matthieu REICHSTADT^{1, 2}, Brigitte PICARD^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Genomics and Computational Biology, doi: 10.18547/gcb.2015.vol1.iss1.e16.– Publié le 06 septembre 2015.

Documents supplémentaires :

1. Method used in GO module script for enrichment calculation for a GO term
2. list of 133 selected protein identifiers in humans to perform the tutorial functionality of ProteINSIDE

Supplemental document 1: Method used in GO module script for enrichment calculation for a GO term.

ProteINSIDE uses a 2x2 contingency table and a Fisher Exact test for equality of proportions. The Fisher Exact test calculates the probability (p) of obtaining a specific GO term given by the hypergeometric distribution as follow:

$$p = \frac{\binom{A+B}{A} \binom{C+D}{C}}{\binom{T}{A+C}} = \frac{T_1! T_2! T'! T''!}{A! B! C! D! T}$$

A is the number of proteins annotated by a specific GO term in the dataset, B the total number of proteins annotated in the dataset, C the number of proteins that are annotated by the specific GO term within the genome (also called “background of a GO term”) and D the total number of annotated proteins in the genome.

	Specific GO term	All GO terms	Total
Annotated proteins in the dataset	A	B	$T_1=A+B$
Annotated proteins in the species	C	D	$T_2=C+D$
Total	$T'=A+C$	$T''=B+D$	$T=T_1+T_2+T'+T''$

Supplemental table 1: List of 133 selected protein identifiers in humans to perform the tutorial functionality of ProteINSIDE.

UniProt_AC	Protein ID	Gene Name
Q15848	ADIPO_HUMAN	ADIPOQ
Q7Z4H4	ADM2_HUMAN	ADM2
P35318	ADML_HUMAN	ADM
Q9BRR6	ADPGK_HUMAN	ADPGK
P04075	ALDOA_HUMAN	ALDOA
P05062	ALDOB_HUMAN	ALDOB
P09972	ALDOC_HUMAN	ALDOC
P16860	ANFB_HUMAN	NPPB
P01160	ANF_HUMAN	NPPA
Q9ULZ1	APEL_HUMAN	APLN
O75947	ATP5H_HUMAN	ATP5H
P06576	ATPB_HUMAN	ATP5B
P0C862	C1T9A_HUMAN	C1QTNF9
P06881	CALCA_HUMAN	CALCA
P10092	CALCB_HUMAN	CALCB
P01258	CALC_HUMAN	CALCA
P06307	CCKN_HUMAN	CCK
A6NKQ9	CGB1_HUMAN	CGB1
Q6NT52	CGB2_HUMAN	CGB2
P01233	CGHB_HUMAN	CGB8
P01189	COL1_HUMAN	POMC
P53621	COPA_HUMAN	COPA
Q5HYK3	COQ5_HUMAN	COQ5
P13073	COX41_HUMAN	COX41
P20674	COX5A_HUMAN	COX5A
P06850	CRF_HUMAN	CRH
P0DML2	CSH1_HUMAN	CSH1
Q14406	CSHL_HUMAN	CSHL1
Q02221	CX6A2_HUMAN	COX6A2
P06733	ENOA_HUMAN	ENO1
P13929	ENOB_HUMAN	ENO3
P09104	ENOG_HUMAN	ENO2
O75356	ENTP5_HUMAN	ENTPD5
P01588	EPO_HUMAN	EPO
P16118	F261_HUMAN	PFKFB1
O60825	F262_HUMAN	PFKFB2
Q16875	F263_HUMAN	PFKFB3
Q16877	F264_HUMAN	PFKFB4
P01225	FSHB_HUMAN	FSHB
O14556	G3PT_HUMAN	GAPDHS
P04406	G3P_HUMAN	GAPDH
P06744	G6PI_HUMAN	GPI
P22466	GALA_HUMAN	GAL
Q9UBC7	GALP_HUMAN	GALP
P01350	GAST_HUMAN	GAST
Q9UBU3	GHRL_HUMAN	GHRL
P09681	GIP_HUMAN	GIP
P01215	GLHA_HUMAN	CGA
P01275	GLUC_HUMAN	GCG
P01148	GON1_HUMAN	GNRH1
O43555	GON2_HUMAN	GNRH2
Q96791	GPHA2_HUMAN	GPHA2
Q86YW7	GPHB5_HUMAN	GPHB5
P81172	HEPC_HUMAN	HAMP
P19367	HXK1_HUMAN	HK1
P52789	HXK2_HUMAN	HK2
P52790	HXK3_HUMAN	HK3
P35557	HXK4_HUMAN	GCK
P10997	IAPP_HUMAN	IAPP
P01344	IGF2_HUMAN	IGF2
P05111	INHA_HUMAN	INHA
P08476	INHBA_HUMAN	INHBA
P09529	INHBB_HUMAN	INHBB
P55103	INHBC_HUMAN	INHBC
P58166	INHBE_HUMAN	INHBE
P51460	INSL3_HUMAN	INSL3
Q14641	INSL4_HUMAN	INSL4
Q9Y5Q6	INSL5_HUMAN	INSL5
Q9Y581	INSL6_HUMAN	INSL6
P01308	INS_HUMAN	INS
Q9UEF7	KLOT_HUMAN	KL
P14618	KPYM_HUMAN	PKM
P30613	KPYR_HUMAN	PKLR

P00338	LDHA_HUMAN	LDHA
P07195	LDHB_HUMAN	LDHB
P07864	LDHC_HUMAN	LDHC
P01229	LSHB_HUMAN	LHB
P20382	MCH_HUMAN	PMCH
P12872	MOTI_HUMAN	MLN
O95299	NDUAA_HUMAN	NDUFA10
O75306	NDUS2_HUMAN	NDUFS2
O75489	NDUS3_HUMAN	NDUFS3
O00217	NDUS8_HUMAN	NDUFS8
P01178	NEU1_HUMAN	OXT
P01185	NEU2_HUMAN	AVP
Q02218	ODO1_HUMAN	OGDH
P10515	ODP2_HUMAN	DLAT
P29803	ODPAT_HUMAN	PDHA2
P08559	ODPA_HUMAN	PDHA1
P11177	ODPB_HUMAN	PDHB
P61366	OSTN_HUMAN	OSTN
P18509	PACA_HUMAN	ADCYAP1
P01298	PAHO_HUMAN	PPY
P17858	PFKAL_HUMAN	PFKL
P08237	PFKAM_HUMAN	PFKM
Q01813	PFKAP_HUMAN	PFKP
P18669	PGAM1_HUMAN	PGAM1
P15259	PGAM2_HUMAN	PGAM2
P00558	PGK1_HUMAN	PGK1
P07205	PGK2_HUMAN	PGK2
P36871	PGM1_HUMAN	PGM1
P07738	PMGE_HUMAN	BPGM
P01236	PRL_HUMAN	PRL
P81277	PRRP_HUMAN	PRLH
P12272	PTHR_HUMAN	PTHLH
P01270	PTHY_HUMAN	PTH
P10082	PYY_HUMAN	PYY
P31930	QCR1_HUMAN	UQCRC1
P04808	REL1_HUMAN	RLN1
P04090	REL2_HUMAN	RLN2
Q8WXF3	REL3_HUMAN	RLN3
Q9BQ08	RETNB_HUMAN	RETNLB
Q9HD89	RETN_HUMAN	RETN
P09683	SECR_HUMAN	SCT
P61278	SMS_HUMAN	SST
P01242	SOM2_HUMAN	GH2
P01241	SOMA_HUMAN	GH1
P52823	STC1_HUMAN	STC1
O76061	STC2_HUMAN	STC2
P01266	THYG_HUMAN	TG
Q8N2E6	TOR2X_HUMAN	TOR2A
P60174	TPI5_HUMAN	TPI1
P40225	TPO_HUMAN	THPO
P20396	TRH_HUMAN	TRH
P01222	TSHB_HUMAN	TSHB
P02766	TTHY_HUMAN	TTR
P55089	UCN1_HUMAN	UCN
Q96RP3	UCN2_HUMAN	UCN2
Q969E3	UCN3_HUMAN	UCN3
Q76510	UTS2B_HUMAN	UTS2B
O95399	UTS2_HUMAN	UTS2
P01282	VIP_HUMAN	VIP

Annexe A.2 : Documents supplémentaires de l'article de Plos ONE

ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses

Nicolas KASPRIC^{1, 2}, Brigitte PICARD^{1, 2}, Matthieu REICHSTADT^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Plos ONE, doi:10.1371/journal.pone.0128086.– Publié le 22 mai 2015.

Documents supplémentaires :

1. List of the 143 proteins from the ontogenesis of adipose tissue (Taga)
2. List of the 13 proteins from the cluster 1 of the ontogenesis of adipose tissue (Taga)
3. List of the 25 proteins from the cluster 8 of the ontogenesis of adipose tissue (Taga)
4. List of the 120 proteins from the ontogenesis of muscle tissue (Chaze)
5. List of the 20 proteins from the cluster 1 of the ontogenesis of muscle tissue (Chaze)
6. List of the 7 proteins from the cluster 8 of the ontogenesis of muscle tissue (Chaze)
7. List of the 46 common proteins from both tissues

Supplemental table 1: List of the 143 proteins from the ontogenesis of adipose tissue.

Uniprot_AC	Protein ID	Gene Name
P31946	1433B_HUMAN	YWHAB
P62258	1433E_HUMAN	YWHAE
P61981	1433G_HUMAN	YWHAG
P31937	3HIDH_HUMAN	HIBADH
P01009	A1AT_HUMAN	SERPINA1
Q96IU4	ABHEB_HUMAN	ABHD14B
Q9UKU7	ACAD8_HUMAN	ACAD8
P49748	ACADV_HUMAN	ACADVL
P07108	ACBP_HUMAN	DBI
Q99798	ACON_HUMAN	ACO2
P63261	ACTG_HUMAN	ACTG1
P61163	ACTZ_HUMAN	ACTR1A
Q15848	ADIPO_HUMAN	ADIPOQ
P55263	ADK_HUMAN	ADK
P49189	AL9A1_HUMAN	ALDH9A1
P02768	ALBU_HUMAN	ALB
P05091	ALDH2_HUMAN	ALDH2
P28838	AMPL_HUMAN	LAP3
P04083	ANXA1_HUMAN	ANXA1
P12429	ANXA3_HUMAN	ANXA3
P09525	ANXA4_HUMAN	ANXA4
P08758	ANXA5_HUMAN	ANXA5
P02647	APOA1_HUMAN	APOA1
P02652	APOA2_HUMAN	APOA2
O75947	ATP5H_HUMAN	ATP5H
P06576	ATPB_HUMAN	ATP5B
P00918	CAH2_HUMAN	CA2
P47756	CAPZB_HUMAN	CAPZB
P52907	CAZA1_HUMAN	CAPZA1
P47755	CAZA2_HUMAN	CAPZA2
P10809	CH60_HUMAN	HSPD1
Q9Y696	CLIC4_HUMAN	CLIC4
P12110	CO6A2_HUMAN	COL6A2
Q5HYK3	COQ5_HUMAN	COQ5
O75208	COQ9_HUMAN	COQ9
P13073	COX41_HUMAN	COX4I1
P20674	COX5A_HUMAN	COX5A
Q02221	CX6A2_HUMAN	COX6A2
P00167	CYB5_HUMAN	CYB5A
P06132	DCUP_HUMAN	UROD
O94760	DDAH1_HUMAN	DDAH1
P09622	DLDH_HUMAN	DLI
Q16555	DPYL2_HUMAN	DPYSL2
Q13011	ECH1_HUMAN	ECH1
P40939	ECHA_HUMAN	HADHA
P30084	ECHM_HUMAN	ECHS1
P49411	EFTU_HUMAN	TUFM
P06733	ENOA_HUMAN	ENO1
P14625	ENPL_HUMAN	HSP90B1
O94905	ERLN2_HUMAN	ERLIN2
P09467	F16P1_HUMAN	FBP1
P15090	FABP4_HUMAN	FABP4
Q96GK7	FAH2A_HUMAN	FAHD2A
P02771	FETA_HUMAN	AFP
P02679	FIBG_HUMAN	FGG
Q02790	FKBP4_HUMAN	FKBP4
P11413	G6PD_HUMAN	G6PD
P51570	GALK1_HUMAN	GALK1
Q14697	GANAB_HUMAN	GANAB
P50395	GDIB_HUMAN	GDI2
P52565	GDIR1_HUMAN	ARHGDI1A
P21695	GPDA_HUMAN	GPD1
P43304	GPDH_HUMAN	GPD2
P38646	GRP75_HUMAN	HSPA9
P68871	HBB_HUMAN	HBB
P13716	HEM2_HUMAN	ALAD
Q6NVY1	HIBCH_HUMAN	HIBCH
P61978	HNRPK_HUMAN	HNRNP1
P04792	HSPB1_HUMAN	HSPB1
O43464	HTRA2_HUMAN	HTRA2
P50213	IDH3A_HUMAN	IDH3A
O75874	IDHC_HUMAN	IDH1
P63241	IF5A1_HUMAN	EIF5A
Q16891	IMMT_HUMAN	IMMT
Q9H2U2	IPYR2_HUMAN	PPA2
P26440	IVD_HUMAN	IVD
P12277	KCRB_HUMAN	CKB
P07195	LDHB_HUMAN	LDHB
P20700	LMNB1_HUMAN	LMNB1
Q03252	LMNB2_HUMAN	LMNB2
P48163	MAOX_HUMAN	ME1
P40925	MDHC_HUMAN	MDH1
Q10713	MPPA_HUMAN	PMPCA
O75431	MTX2_HUMAN	MTX2
Q14764	MVP_HUMAN	MVP
O95299	NDUAA_HUMAN	NDUFA10
P28331	NDUS1_HUMAN	NDUFS1

O75306	NDUS2_HUMAN	NDUFS2
O75489	NDUS3_HUMAN	NDUFS3
O00217	NDUS8_HUMAN	NDUFS8
P19404	NDUV2_HUMAN	NDUFV2
P12694	ODBA_HUMAN	BCKDHA
Q02218	ODO1_HUMAN	OGDH
P36957	ODO2_HUMAN	DLST
P10515	ODP2_HUMAN	DLAT
P11177	ODPB_HUMAN	PDHB
Q92882	OSTF1_HUMAN	OSTF1
Q99497	PARK7_HUMAN	PARK7
P05166	PCCB_HUMAN	PCCB
P30101	PDIA3_HUMAN	PDIA3
P18669	PGAM1_HUMAN	PGAM1
P35232	PHB_HUMAN	PHB
O60240	PLIN1_HUMAN	PLIN1
P00491	PNPH_HUMAN	PNP
P32119	PRDX2_HUMAN	PRDX2
P30048	PRDX3_HUMAN	PRDX3
P30044	PRDX5_HUMAN	PRDX5
P30041	PRDX6_HUMAN	PRDX6
P25786	PSA1_HUMAN	PSMA1
P25788	PSA3_HUMAN	PSMA3
Q06323	PSME1_HUMAN	PSME1
Q9UL46	PSME2_HUMAN	PSME2
Q8N8N7	PTGR2_HUMAN	PTGR2
P11498	PYC_HUMAN	PC
P31930	QCR1_HUMAN	UQCRC1
Q15293	RCN1_HUMAN	RCN1
P08865	RSSA_HUMAN	RPSA
Q13228	SBP1_HUMAN	SELENBP1
P31040	SDHA_HUMAN	SDHA
Q15019	SEPT2_HUMAN	SEPT2
P00441	SODC_HUMAN	SOD1
P04179	SODM_HUMAN	SOD2
P31948	STIP1_HUMAN	STIP1
Q9Y3F4	STRAP_HUMAN	STRAP
Q9P2R7	SUCB1_HUMAN	SUCLA2
P49588	SYAC_HUMAN	AARS
P37837	TALDO_HUMAN	TALDO1
P07437	TBB5_HUMAN	TUBB
P17987	TCPA_HUMAN	TCP1
P55072	TERA_HUMAN	VCP
Q5T1C6	THEM4_HUMAN	THEM4
Q9BWD1	THIC_HUMAN	ACAT2
P25325	THTM_HUMAN	MPST
Q9NYL9	TMOD3_HUMAN	TMOD3
P60174	TPI5_HUMAN	TPI1
P09493	TPM1_HUMAN	TPM1
P06753	TPM3_HUMAN	TPM3
P51580	TPMT_HUMAN	TPMT
P02787	TRFE_HUMAN	TF
P02766	TTHY_HUMAN	TTR
P45880	VDAC2_HUMAN	VDAC2
P08670	VIME_HUMAN	VIM
P18206	VINC_HUMAN	VCL

Supplemental table 2: List of the 13 proteins from the cluster 1 of the ontogenesis of adipose tissue.

UniProt ID	UniProt AC	Gene Name
ALBU_HUMAN	P02768	ALB
TPM3_HUMAN	P06753	TPM3
VIME_HUMAN	P08670	VIM
TCPA_HUMAN	P17987	TCP1
LMNB1_HUMAN	P20700	LMNB1
CAPZB_HUMAN	P47756	CAPZB
SYAC_HUMAN	P49588	AARS
CAZA1_HUMAN	P52907	CAPZA1
HNRPK_HUMAN	P61978	HNRNPK
IF5A1_HUMAN	P63241	EIF5A
LMNB2_HUMAN	Q03252	LMNB2
TMOD3_HUMAN	Q9NYL9	TMOD3
STRAP_HUMAN	Q9Y3F4	STRAP

Supplemental table 3: List of the 25 proteins from the cluster 8 of the ontogenesis of adipose tissue.

UniProt ID	UniProt AC	Gene Name
MTX2_HUMAN	O75431	MTX2
CAH2_HUMAN	P00918	CA2
ALBU_HUMAN	P02768	ALB
SODM_HUMAN	P04179	SOD2
PCCB_HUMAN	P05166	PCCB
F16P1_HUMAN	P09467	FBP1
DLDH_HUMAN	P09622	DLD
HEM2_HUMAN	P13716	ALAD
PGAM1_HUMAN	P18669	PGAM1
GPDA_HUMAN	P21695	GPD1
ECHM_HUMAN	P30084	ECHS1
QCR1_HUMAN	P31930	UQCRC1
ODO2_HUMAN	P36957	DLST
VDAC2_HUMAN	P45880	VDAC2
AL9A1_HUMAN	P49189	ALDH9A1
IDH3A_HUMAN	P50213	IDH3A
TPIS_HUMAN	P60174	TPI1
HBB_HUMAN	P68871	HBB
ODO1_HUMAN	Q02218	OGDH
ECH1_HUMAN	Q13011	ECH1
THEM4_HUMAN	Q5T1C6	THEM4
HIBCH_HUMAN	Q6NVY1	HIBCH
PARK7_HUMAN	Q99497	PARK7
THIC_HUMAN	Q9BWD1	ACAT2
IPYR2_HUMAN	Q9H2U2	PPA2

Supplemental table 4: List of the 120 proteins from the ontogenesis of muscle tissue.

UniProt_AC	Protein ID	Gene Name
Q15019	SEPT2_HUMAN	SEPT2
Q9NVA2	SEP11_HUMAN	SPET11
P68133	ACTS_HUMAN	ACTA1
P68032	ACTC_HUMAN	ACTC1
Q9P1U1	ARP3B_HUMAN	ACTR3B
Q15848	ADIPO_HUMAN	ADIPOQ
Q8NDY3	ARHL1_HUMAN	ADPRHL1
P02771	FETA_HUMAN	AFP
O43865	SAHH2_HUMAN	AHCYL1
O95433	AHSA1_HUMAN	AHSA1
O00170	AIP_HUMAN	AIP
O60218	AK1BA_HUMAN	AKR1B10
P02768	ALBU_HUMAN	ALB
P49419	AL7A1_HUMAN	ALDH7A1
P49189	AL9A1_HUMAN	ALDH9A1
P04083	ANXA1_HUMAN	ANXA1
P07355	ANXA2_HUMAN	ANXA2
P09525	ANXA4_HUMAN	ANXA4
P08758	ANXA5_HUMAN	ANXA5
P02647	APOA1_HUMAN	APOA1
Q9Y235	ABEC2_HUMAN	ABOEC2
P52565	GDIR1_HUMAN	ARHGDI1A
O75947	ATP5H_HUMAN	ATP5H
P47755	CAZA2_HUMAN	CAPZA2
P32239	GASR_HUMAN	CCKBR
P78371	TCPB_HUMAN	CCT2
P48643	TCPE_HUMAN	CCT5
Q9Y281	COF2_HUMAN	CFL2
P06732	KCRM_HUMAN	CKM
O00299	CLIC1_HUMAN	CLIC1
Q9Y696	CLIC4_HUMAN	CLIC4
Q15417	CNN3_HUMAN	CNN3
Q98T78	CSN4_HUMAN	COPS4
P11182	ODB2_HUMAN	DBT
O94760	DDAH1_HUMAN	DDAH1
O95865	DDAH2_HUMAN	DDAH2
Q13838	DX39B_HUMAN	DDX39B
Q6ICB0	DES1_HUMAN	DES1
Q13409	DC1I2_HUMAN	DYNC1I2
Q13347	EIF3I_HUMAN	EIF3I
Q14240	IF4A2_HUMAN	EIF4A2
P63241	IF5A1_HUMAN	EIF5A
P06733	ENOA_HUMAN	ENO1
P13929	ENOB_HUMAN	ENO3
P13726	TF_HUMAN	F3
Q96GK7	FAH2A_HUMAN	FAHD2A
Q8TBF2	PGFS_HUMAN	FAM213B
Q96C11	FGGY_HUMAN	FGGY
Q02790	FKBP4_HUMAN	FKBP4
Q16658	FSCN1_HUMAN	FSCN1
Q14353	GAMT_HUMAN	GAMT
P41250	SYG_HUMAN	GARS
P50395	GDIB_HUMAN	GDI2
P06396	GELS_HUMAN	GSN
Q9H0R4	HDHD2_HUMAN	HDHD2
P31942	HNRH3_HUMAN	HNRNPH3
P61978	HNRPK_HUMAN	HNRNPK
P11142	HSP7C_HUMAN	HSPA8
P38646	GRP75_HUMAN	HSPA9
P04792	HSPB1_HUMAN	HSPB1
O14558	HSPB6_HUMAN	HSPB6
P50213	IDH3A_HUMAN	IDH3A
Q9NPH2	INO1_HUMAN	ISYNA1
P07195	LDHB_HUMAN	LDHB
Q96DT0	LEG12_HUMAN	LGALS12
P02545	LMNA_HUMAN	LMNA
P26038	MOES_HUMAN	MSN
P56278	MTCP1_HUMAN	MTCP1
Q13203	MYBPH_HUMAN	MYBPH
P01106	MYC_HUMAN	MYC
P12882	MYH1_HUMAN	MYH1
P05976	MYL1_HUMAN	MYL1
P10916	MLRV_HUMAN	MYL2
P12829	MYL4_HUMAN	MYL4
P14649	MYL6B_HUMAN	MYL6B
P24844	MYL9_HUMAN	MYL9
Q96A32	MLRS_HUMAN	MYLPF
P54920	SNAA_HUMAN	NAPA
P28331	NDUS1_HUMAN	NDUFS1
P07237	PDIA1_HUMAN	P4HB
Q9UQ80	PA2G4_HUMAN	PA2G4
P68402	PA1B2_HUMAN	PAFAH1B2
Q99497	PARK7_HUMAN	PARK7
P30101	PDIA3_HUMAN	PDIA3
P35232	PHB_HUMAN	PHB
Q8N8W4	PLPL1_HUMAN	PNPLA1
Q08752	PPID_HUMAN	PPID
P62136	PP1A_HUMAN	PPP1CA
P62140	PP1B_HUMAN	PPP1CB
P32119	PRDX2_HUMAN	PRDX2
P30041	PRDX6_HUMAN	PRDX6
P10644	KAPO_HUMAN	PRKAR1A
Q9UMS4	PRP19_HUMAN	PRPF19
P25786	PSA1_HUMAN	PSMA1
P60900	PSA6_HUMAN	PSMA6
Q99436	PSB7_HUMAN	PSMB7
P35998	PRS7_HUMAN	PSMC2
Q06323	PSME1_HUMAN	PSME1
Q9UL46	PSME2_HUMAN	PSME2
Q8N8N7	PTGR2_HUMAN	PTGR2
Q8NHV5	RLAQL_HUMAN	RPLP0P6
Q9Y230	RUVB2_HUMAN	RUVBL2
Q13228	SBP1_HUMAN	SELENBP1
P49903	SPS1_HUMAN	SEPHS1
P01009	A1AT_HUMAN	SERPINA1
Q9UNH7	SNX6_HUMAN	SNX6
P31948	STIP1_HUMAN	STIP1
P16949	STMN1_HUMAN	STMN1
Q9P2R7	SUCB1_HUMAN	SUCLA2
P17987	TCPA_HUMAN	TCP1
P13805	TNNT1_HUMAN	TNNT1
P45378	TNNT3_HUMAN	TNNT3
P60174	TPI5_HUMAN	TPI1
P49411	EFTU_HUMAN	TUFM
Q8TBC4	UBA3_HUMAN	UBA3
P31930	QCR1_HUMAN	UQCRC1
P23381	SYWC_HUMAN	WARS
O75083	WDR1_HUMAN	WDR1
P62258	1433E_HUMAN	YWHAE
P63104	1433Z_HUMAN	YWHAZ

Supplemental table 5: List of the 20 proteins from the cluster 1 of the ontogenesis of muscle tissue.

UniProt ID	UniProt AC	Gene Name
KAPO_BOVIN	P00514	PRKAR1A
AT5F1_BOVIN	P13620	ATP5H
Q3ZC09_BOVIN	P15429	ENO3
EFTU_BOVIN	P49410	TUFM
PNPH_BOVIN	P55859	PNP / NP
/	P56965	DDAH1
ANXA5_BOVIN	P81287	ANXA5 / ANX5
SUCB2_BOVIN	Q148D5	SUCLA2
IDH3B_BOVIN	Q148J8	IDH3A
AL7A1_BOVIN	Q2KJC9	ALDH7A1
AL9A1_BOVIN	Q2KJH9	ALDH9A1
/	Q2NKY7	SEPT2
ALBU_BOVIN	Q3I349	ALB
Q3MVK9_BOVIN	Q3MVK9	FSCN1
/	Q3MHL4	AHCY
/	Q3SZ65	EIF4A2
Q3ZBM1_BOVIN	Q3ZBM1	ADPRHL1 / ARH2
/	Q5E966	EIF3I / EIF3S2
PRDX2_BOVIN	Q9BGI3	PRDX2
ENOA_BOVIN	Q9XSJ4	ENO1

Supplemental table 6 : List of the 7 proteins from the cluster 8 of the ontogenesis of muscle tissue.

UniProt ID	UniProt AC	Gene Name
/	P02769	ALB
PDIA1_BOVIN	P05307	PDIA1
/	Q0VBZ1	MYBPH
Q2TBP0_BOVIN	Q2TBP0	PSMB7
/	Q3ZBH0	CCT2
ACTC_BOVIN	Q3ZC07	ACTC1 / ACTC
ENOA_BOVIN	Q9XSJ4	ENO1

Supplemental table 8 : List of the 46 common proteins from both tissues.

UniProt_AC	Protein ID	Gene Name
P62258	1433E_HUMAN	YWHAE
P01009	A1AT_HUMAN	SERPINA1
Q15848	ADIPO_HUMAN	ADIPOQ
P49189	AL9A1_HUMAN	ALDH9A1
P02768	ALBU_HUMAN	ALB
P04083	ANXA1_HUMAN	ANXA1
P09525	ANXA4_HUMAN	ANXA4
P08758	ANXA5_HUMAN	ANXA5
P02647	APOA1_HUMAN	APOA1
O75947	ATP5H_HUMAN	ATP5H
P47755	CAZA2_HUMAN	CAPZA2
Q9Y696	CLIC4_HUMAN	CLIC4
O94760	DDAH1_HUMAN	DDAH1
P49411	EFTU_HUMAN	TUFM
P06733	ENOA_HUMAN	ENO1
P02771	FETA_HUMAN	AFP
Q96C11	FGGY_HUMAN	FGGY
Q02790	FKBP4_HUMAN	FKBP4
P32239	GASR_HUMAN	CCKBR
P50395	GDIB_HUMAN	GDI2
P52565	GDIR1_HUMAN	ARHGDI1
P38646	GRP75_HUMAN	HSPA9
P61978	HNRPK_HUMAN	HNRNPK
P04792	HSPB1_HUMAN	HSPB1
P50213	IDH3A_HUMAN	IDH3A
P63241	IF5A1_HUMAN	EIF5A
P07195	LDHB_HUMAN	LDHB
P56278	MTCP1_HUMAN	MTCP1
P28331	NDUS1_HUMAN	NDUFS1
Q99497	PARK7_HUMAN	PARK7
P30101	PDIA3_HUMAN	PDIA3
P35232	PHB_HUMAN	PHB
Q8N8W4	PLPL1_HUMAN	PNPLA1
P32119	PRDX2_HUMAN	PRDX2
P30041	PRDX6_HUMAN	PRDX6
P25786	PSA1_HUMAN	PSMA1
Q06323	PSME1_HUMAN	PSME1
Q9UL46	PSME2_HUMAN	PSME2
Q8N8N7	PTGR2_HUMAN	PTGR2
P31930	QCR1_HUMAN	UQCRC1
Q13228	SBP1_HUMAN	SELENBP1
Q15019	SEPT2_HUMAN	SEPT2
P31948	STIP1_HUMAN	STIP1
Q9P2R7	SUCB1_HUMAN	SUCLA2
P13726	TF_HUMAN	F3
P60174	TPIS_HUMAN	TPI1

Annexe A.3 : Article**Quest for biomarkers of the lean-to-fat ratio by Proteomics in beef production**

Muriel BONNET^{1,2}, Nicolas KASPRIC^{1,2} and Brigitte PICARD^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Proceeding of the 3rd meeting of working groups 1, 2, 3 of COST Action FA1002, Kosice, Slovaquie, April 25-26 2013.

Quest for biomarkers of the lean-to-fat ratio by proteomics in beef production

Muriel Bonnet^{1,2}, Nicolas Kaspric^{1,2}, Brigitte Picard^{1,2}

¹INRA, UMR1213 Herbivores, 63122 Saint-Genès-Champagnelle, France;

muriel.bonnet@clermont.inra.fr

²VetAgro Sup, Élevage et production des ruminants, 63370 Lempdes, France

Producing meat animals with adequate muscular and adipose masses (i.e. lean-to-fat ratio) is an economic challenge for the beef industry. In cattle, conformation (muscle mass) and fatness (adipose tissue mass) are evaluated using the European Union beef carcass classification system (EUROP) scale. The EUROP scale determines the price/kg of carcasses. The lean-to-fat ratio is the result of a dynamic balance between the number and size of muscular and adipose cells, respectively (Bonnet *et al.*, 2010). Identifying proteins that contribute to the increase in the number and volume of adipose and muscular cells has implication for the proposition of biomarkers of growth potential and/or of carcass composition.

The total number of muscle fibres is set by the end of the second trimester of gestation (Picard *et al.*, 2002). Conversely, the number of adipocytes is set by birth or by early adulthood, depending on the anatomical location of the adipose tissue (Vernon, 1986). Thus, we hypothesized that a high throughput molecular characterisation of adipose tissue (AT) and muscle from bovine foetuses differing by the age would be a powerful way to identify proteins associated to the increase in the number and the size of muscular and adipose cells.

We combined measurements of chemical composition, cellularity, histology, enzyme activities, gene expression and proteomics to describe the ontogeny of perirenal AT and *Semitendinosus* muscle in bovine at 60, 110, 180, 210 and 260 days post conception (dpc) in Blond d'Aquitaine (n=3 per age) and Charolais (n=5 per age) breeds. These breeds were chosen for their differences in lean-to-fat ratio in the post-natal life.

Between 110 and 260 dpc (38 and 90% of gestation length, respectively), the increase in the weight of perirenal AT resulted from an increase in the volume and mainly in the number of adipocytes (Taga *et al.*, 2011). The increases in adipocyte volume and number were accompanied by changes in the abundance of 128 proteins among the 143 proteins identified and common to the four last fetal ages studied (Taga *et al.*, 2012). Among the identified proteins, some of them have never been described in the AT and may contribute to hyperplasia of adipose precursors, by controlling cell cycle progression, apoptosis and/or by delaying adipocyte differentiation. The age of 180 dpc seems to be a pivotal age for the transition between proliferation and differentiation of adipocyte progenitors. An increase in the abundance of many proteins involved in differentiation and in the increase in adipocyte volume was observed from 180 dpc.

The fetal growth of semitendinosus muscle was accompanied by changes in the abundance of 245 proteins. We revealed high abundance of proteins involved in apoptosis at 60 and 110 dpc, suggesting that the proliferation – apoptosis balance may play a role in the determination of the total number of fibers (Chaze *et al.*, 2008). The age of 180 dpc corresponds to a reduction in cell proliferation and a transition between the formation of myofibres and their maturation. Indeed, the abundance of stathmin, which has an important role in cell cycle regulation, decreased from 180 dpc onwards. Septin proteins (septin 2 and 11 isoforms) involved in cytoskeletal organisation, scaffolding and cell division plane had a stable expression up to 180 dpc, consistent with intense cell division until this stage, and declined thereafter. Annexin A1, which has an anti proliferative function via the activation of ERK pathways, showed increased abundance from 180 dpc onwards. Increased maturation of fibers from 180 to 260 dpc is reflected by significant changes in the profiles of protein isoforms belonging to metabolic and contractile pathways (Chaze *et al.*, 2009).

The cellular and molecular features of AT and muscle during ontogenesis provide potential hallmarks of adipose and muscular cells hyperplasia or hypertrophy. Among these proteins we hypothesize there are ‘master’ proteins, thus a current integrative bioinformatic analysis of adipose and muscular data aims to identify them. Then, the relationship between the abundance of these proteins and data from carcass composition will be studied in tissues from meat producing cattle differing by their rearing conditions or by their genotypes.

References

- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B., 2010. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* 4: 1093-1109.
- Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B., 2008. *In vivo* proteome dynamics during early bovine myogenesis. *Proteomics* 8: 4236-4248.
- Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B., 2009. Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle. *Animal* 3: 980-1000.
- Picard, B., Lefaucheur, L., Berri, C. and Duclos, M.J., 2002. Muscle fibre ontogenesis in farm animal species. *Reproduction, Nutrition, Development* 42: 415-431.
- Taga, H., Bonnet, M., Picard, B., Zingaretti, M.C., Cassar-Malek, I., Cinti, S. and Chilliard, Y., 2011. Adipocyte metabolism and cellularity are related to differences in adipose tissue maturity between Holstein and Charolais or Blond d'Aquitaine fetuses. *Journal of Animal Science* 89: 711-721.
- Taga, H., Chilliard, Y., Meunier, B., Chambon, C., Picard, B., Zingaretti, M.C., Cinti, S. and Bonnet, M., 2012. Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown? *Journal of Cellular Physiology* 227: 1688-1700.
- Vernon, R.G., 1986. The growth and metabolism of Adipocytes. In: P.J. Buttery, D.B. Lindsay and N.B. Haynes (Eds.), *Control and Manipulation of Animal Growth*. Butterworths, London., pp. 67-83.

Annexe A.4 : Résumé et présentation affichée

ProteINSIDE: a web service to computerize an in-depth analysis of functions, sequences, secretions and interactions for proteins

Nicolas KASPRIC^{1, 2}, Matthieu REICHSTADT^{1, 2}, Brigitte PICARD^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

RECOMB Comparative Genomics (Research in Computational Molecular Biology), Lyon, France, October 17-18-19 2013.

ProteINSIDE: a web service to computerize an in-depth analysis of functions, sequences, secretions and interactions for proteins.

KASPRIC Nicolas^{1,2}, REICHSTADT Matthieu^{1,2}, PICARD Brigitte^{1,2}, TOURNAYRE Jérémy^{1,2} and BONNET Muriel^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Corresponding author: nicolas.kaspric@clermont.inra.fr

Given the increasing quantity of genomic data produced, a strategy to perform a systemic and integrative analysis of protein biological meanings is to develop an online workflow with an interface devoted to reachable and fully customizable analysis and to an easy view of the results.

From a list of proteins/genes, ProteINSIDE collects and stores accession numbers, protein sequences and biological meanings from UniprotKB database and NCBI. ProteINSIDE annotates proteins/genes according to the Gene Ontology (GO) from UniprotKB database and calculates over- and under-represented terms. ProteINSIDE proposes a list of putative secreted proteins by both the prediction of signal peptides with SignalP software and the search for GO terms relevant to the secretion process. Protein-protein interactions (PPi) are identified with PsicQuick web service; our script checks 28 PPi databases and collects experimentally proven PPi. PPi are visualized by a cytoscape web implementation, with customizable view and analysis of network based on betweenness and closeness centralities. A web interface for ProteINSIDE gives results by modules of analysis as dynamic tables and diagrams.

We validated this web service on a test list made of human proteins from glycolysis, citric acid cycle and hormones. ProteINSIDE collected sequences and biological meanings (functions, tissue specificity and cellular location) for 100% of these proteins. ProteINSIDE detected 93% of expected signal peptides and annotated at least 100% of the proteins involved in glycolysis and citric acid cycle with GOs related to these pathways.

In conclusion ProteINSIDE saves time and treatment information for researchers who get results from several softwares and database with a single query. ProteINSIDE imports accession numbers (usable in other web services) and biological meanings, identifies over- and under-represented biological functions and secreted proteins, and constructs a network from large lists of proteins/genes or accession numbers. ProteINSIDE is freely available at: <http://147.99.129.193/i-analyse/index.php>

ProteINSIDE: a web service to computerize an in-depth analysis of functions, sequences, secretions and interactions for proteins.

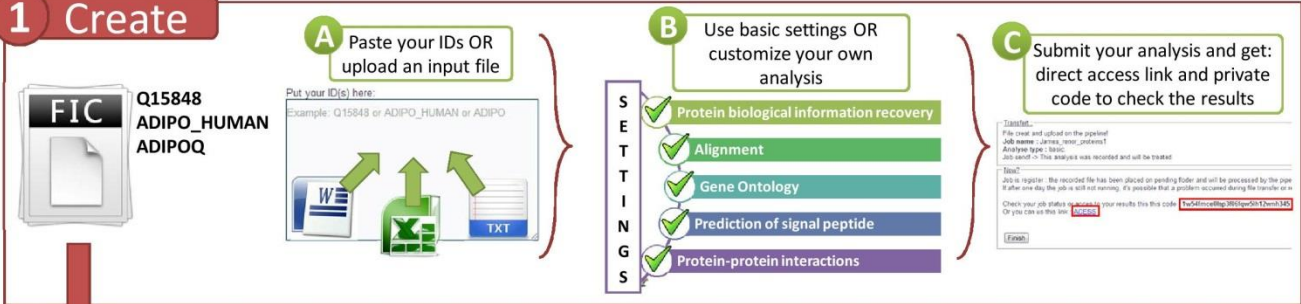
KASPRIK Nicolas^{1,2}, REICHSTADT Matthieu^{1,2}, PICARD Brigitte^{1,2}, TOURNAYRE Jérémy^{1,2} and BONNET Muriel^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

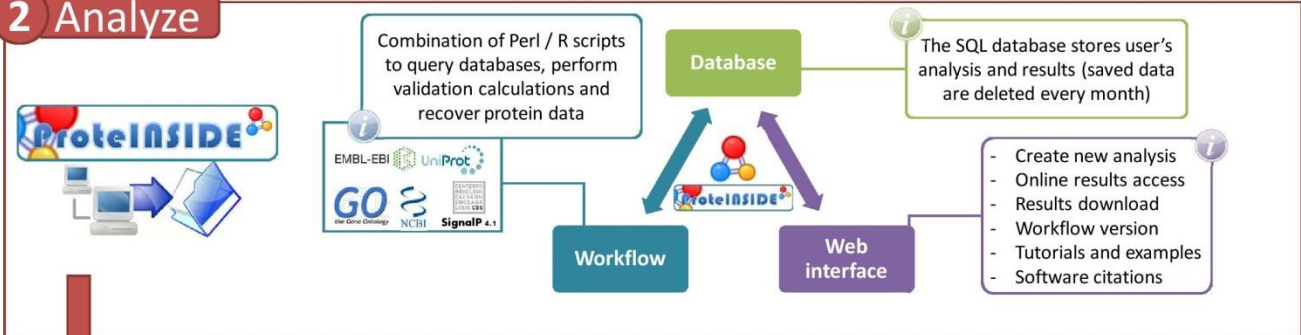
² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Given the increasing quantity of genomic data produced, a strategy to perform a systemic and integrative analysis of protein biological meanings is to develop an online workflow with an interface devoted to reachable and fully customizable analysis and to an easy view of the results.

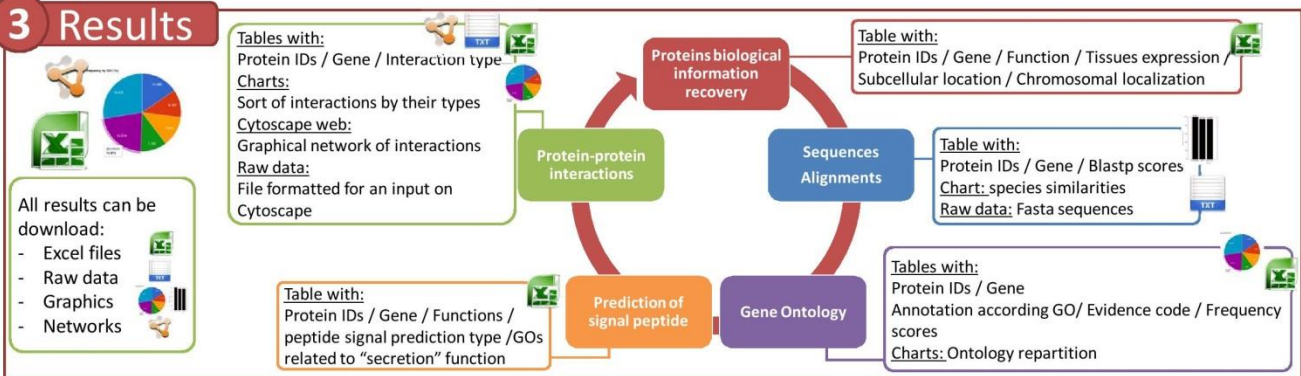
1 Create



2 Analyze



3 Results



ProteINSIDE is a web service to save time and treatment information for researchers who get results from several softwares and databases with a single query. ProteINSIDE imports accession numbers (usable in other web services) and biological meanings, identifies over- and under-represented biological functions and secreted proteins, and constructs a network from large lists of proteins/genes identifiers.

Annexe A.5 : Article**Fat&MuscleDB: Integrating "omics" data from adipose tissue and muscle**

Jérémy TOURNAYRE^{1, 2}, Isabelle CASSAR-MALEK^{1, 2}, Matthieu REICHSTADT^{1, 2}, Brigitte PICARD^{1, 2}, Nicolas KASPRIC^{1,2}, and Muriel BONNET^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Proceeding of the 4rd meeting of working groups 1, 2, 3 of COST Action FA1002, Milano, Italie, November 16-18 2014.

Fat&MuscleDB: Integrating "omics" data from adipose tissue and muscle

Jérémy Tournayre^{1,2}, Isabelle Cassar-Malek^{1,2}, Matthieu Reichstadt^{1,2}, Brigitte Picard^{1,2}, Nicolas Kaspric^{1,2} and Muriel Bonnet^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, ClermontFerrand, France

Corresponding authors: jeremy.tournayre@clermont.inra.fr and muriel.bonnet@clermont.inra.fr

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

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

Methods

A database of transcripts and proteins (with their abundance values) dealing with the growth of adipose and muscle tissue has been built from Pubmed and Gene Expression Omnibus (GEO) on the National Center for Biotechnology Information (NCBI) and the Web of Knowledge. To select contextualized articles or dataset, keywords were defined by experts in fat and muscle growth and submitted via an automatic procedure. Each retrieved publication (with its supplementary data) was analysed manually to check for results and table(s) reporting transcripts or proteins differentially abundant during muscle and/or fat growth. The extraction of tables from a PDF was simplified by the use of a tool that has been implemented. We used "pdftotext" command to convert PDF to text and then to a tabbed file. Transcriptomics results of GEO were analysed to determine transcripts differentially expressed or not during muscle and/or fat growth. Curators had chosen, among a GEO Series

(GSE), GEO Samples (GSM) to submit to analyse in the context of adipose tissue and muscle growths. Transcriptomics results were then recovered and processed using a tool developed in R software with libraries GEOquery (Davis and Meltzer, 2007) and Anapuce (<http://cran.rproject.org/web/packages/anapuce/>). Anapuce was chosen for its specific functions ("DiffAnalysis" and "DiffAnalysis.unpaired") for the differential analysis between two groups of microarray data. Data were normalized when necessary through "normalisation" function, spots for a same transcript and technical replicates were averaged and the variance mixture was calculated with "DiffAnalysis" or "DiffAnalysis.unpaired" functions for paired data or unpaired data, respectively. We chose a p-value of 0.15 for the variance because of the stringency of this test that strongly reduced the number of regulated transcripts. Each of the data analysed by these methods was verified by curators and annotated as "Analysed by Fat&MuscleDB". In contrast, data recorded as they were published were annotated as "Not analysed by Fat&MuscleDB". Data were recorded with information given by the authors regarding the experimental protocols in order to allow investigators to conduct researches and aggregations in Fat&MuscleDB.

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

Results and discussion

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

We report the example of the integration of data related to the criterion "Muscular hypertrophy from genetic origin" (Figure 1). By clicking on the "Aggregation" button, investigators have to select this criterion among others and to select other settings to submit the integration. Currently this integration uses data from six references analysed by twelve comparisons. Thus, UniProt accession lists are built up of transcripts and proteins according to their abundance. After the integration of the data by the aggregation process we identified 13 proteins and 357 transcripts with a decreased abundance, 3 proteins and 286 transcripts with an increased abundance and 17,088 transcripts with a stable abundance that are related to

"Muscular hypertrophy from genetic origin". This list will be submitted for data mining using informatics resources such as ProteINSIDE that we have developed (see the text by N. Kaspric), to identify central or relevant genes or proteins related to muscular hypertrophy. Aggregation can be achieved for the 72 other criteria that we have defined to depict all the stages of adipose tissue and muscle growth.

Decreased abundance	Increased abundance	Stable abundance			
Analysed by Fat&MuscleDB	Uniprot	Entry Name	Reviewed	Number	
no	Q3T149	HSPB1_BOVIN	yes	1	
no	F1N789	F1N789_BOVIN	no	1	
no	A5PJM2	A5PJM2_BOVIN	no	1	
no	A6QPA7	A6QPA7_BOVIN	no	1	
no	P20004	ACON_BOVIN	yes	1	
no	P79334	PYGM_BOVIN	yes	1	
no	O77834	PRDX6_BOVIN	yes	1	
no	Q5E997	CAZA2_BOVIN	yes	1	
no	G3N3C9	G3N3C9_BOVIN	no	1	
no	Q8MKH6	TNNT1_BOVIN	yes	1	
no	A8KC82	A8KC82_BOVIN	no	1	
no	A6QLL8	A6QLL8_BOVIN	no	1	
no	Q32KV0	PGAM2_BOVIN	yes	1	

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

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

Acknowledgements

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

References

- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B., 2010. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* 4: 1093-1109.
- Davis, S. and Meltzer, P.S., 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23: 1846-1847.
- Picard, B., Cassar-Malek, I., Guillemain, N. and Bonnet, M., 2011. Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: M. Moo-Young (Ed.), *Comprehensive Biotechnology (Second Edition)*. Academic Press, Burlington, pp. 395-405.
- Romacho, T., Elsen, M., Rohrborn, D. and Eckel, J., 2014. Adipose tissue and its role in organ crosstalk. *Acta Physiol (Oxf)* 210: 733-753.

Annexe A.6 : Article et présentation affichée

Proteomics data from ruminants easily investigated using
ProteINSIDE

Nicolas KASPRIC^{1, 2}, Brigitte PICARD^{1, 2}, Matthieu REICHSTADT^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Proceeding of the 4rd meeting of working groups 1, 2, 3 of COST Action FA1002, Milano, Italie, November 16-18 2014.

Proteomics data from ruminants easily investigated using ProteINSIDE

Nicolas Kaspric^{1,2}, Brigitte Picard^{1,2}, Matthieu Reichstadt^{1,2}, Jérémy Tournayre^{1,2} and Muriel Bonnet^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Corresponding authors: nicolas.kaspric@clermont.inra.fr and muriel.bonnet@clermont.inra.fr

Objectives

A main challenge for scientists working on the efficiency of ruminant production and the quality of their products (meat, milk...) is to understand which genes and proteins control nutrient metabolism and partitioning between tissues, or which genes and proteins control tissues growth and physiology (Bonnet *et al.*, 2010). Their researches have produced huge amounts of genomic data requiring a lot of bioinformatics analyses to extract meaningful biological context for proteins or genes in ruminants. A strategy to increase the efficiency and the robustness of data mining from ruminant genomics data is to develop an online workflow that integrates several analysis steps in one package as it has been partly done in Human (BioMyn web service (Ramirez *et al.*, 2012)). Here, we present ProteINSIDE, an online workflow to analyse lists of proteins or genes from ruminant species by gathering biological information provided by an overview of data from myriad of databases, annotations according to Gene Ontology (GO), the prediction of tissue secretome and proteins interactions networks.

Methods

ProteINSIDE is an online tool, freely available at www.proteinside.org by using an internet browser. A flow chart (Figure 1) details the proposed basic or customizable analyses and the four main modules of the workflow. Whatever the type of analysis, the workflow uses data from the input file and runs default scripts (basic analysis) or scripts according to settings selected by the user (customs analysis). The four modules of analysis were developed as follow:

- “Identifier mapping” module is an overview of available biological information thank to an assembly of reviewed biological data from UniProt and NCBI databases, and available on the “ID resume” web page
- “Gene Ontology” module queries QuickGO (Huntley *et al.*, 2009) database, with an option to unselect (basic analysis) or to select (custom analysis) electronic annotations. Then it ranks the over- represented GO terms by comparing the frequencies of a GO within a dataset and within the genome by a Fisher exact test. The most linked GO terms are viewed as a GOTree chart that is an ordered tree layout network provided by the custom analysis
- “Secreted proteins” module searches for signal peptide on protein’s fasta sequences thanks to SignalP (Petersen *et al.*, 2011) and TargetP (Emanuelsson *et al.*, 2000) algorithms. To support this prediction, ProteINSIDE selects GO terms related to secretion processes. The GO terms are also analysed to predict proteins that are secreted by processes that do not involve signal peptides (Nickel, 2003)
- “Protein-protein interactions (PPi)” module searches for PPi recorded among the 26 PPi databases of Psicquic (Aranda *et al.*, 2011). ProteINSIDE only imports reviewed and curated PPi, and constructs PPi networks using an online cytoscape view (Lopes *et al.*, 2010) either between proteins of the dataset or between proteins of the dataset and outside of the dataset (in same species or using orthologous gene products)

All results from ProteINSIDE's modules are viewed on the web page or can be downloaded. The output files are Excel file (.xls), Cytoscape file (.png or .pdf or .graphml or .xgmml), text or FASTA file (.txt or .fa) and pictures (.jpg or .png or .pdf).

ProteINSIDE inputs are proteins or genes identifiers (IDs) or names (e.g. ADIPO or ADIPO_HUMAN or e.g. gi|62022275) or UniProt protein accession number (e.g. Q15848) from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). To test ProteINSIDE's performances, we used a dataset composed of 133 proteins: 34 proteins related to the glycolysis cycle, 11 proteins from the respiratory chain, 5 proteins from the tricarboxylic acid cycle, 79 hormones or secreted proteins and proteins with very specific functions unrelated to the others. We also included a duplicated protein among proteins of the glycolysis to verify its recognition by ProteINSIDE. We created this dataset on Bovine species, but the numbers of annotations and PPI weren't sufficient for a clear evaluation of the functionalities of ProteINSIDE. Then, we used the Bovine IDs for an analysis that use knowledge available in Human and ProteINSIDE automatically converted IDs from Bovine to Human using known orthologous gene products.

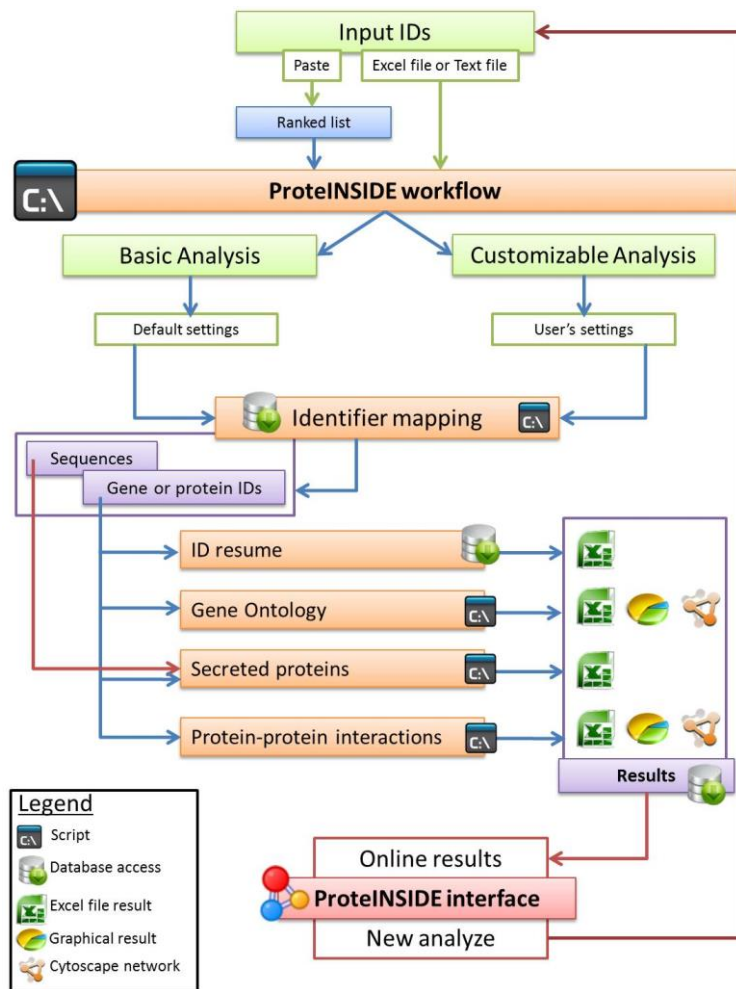


Figure 1. Flow chart of ProteINSIDE structure. The four modules for querying the available biological information, annotations according to the gene ontology, predictions of secreted proteins and protein-protein interactions are either all selected in the basic analysis or individually selected with specific settings in the custom analysis.

Results and discussion

Results from the “Basic Analysis” shown that 133 proteins were recognized by ProteINSIDE, the protein in duplicate was identified and excluded from the analysis. Thus, 132 proteins were submitted to the analyses.

The ID Resume module of ProteINSIDE extracted and summarized biological information about each protein of the dataset. Results are available on the “ID Resume” of the tool bar menu and are summarized in a table with: proteins and genes IDs, biological function, subcellular location, tissue specificity and chromosome location of the gene.

Among the 132 proteins submitted, ProteINSIDE annotates 123 proteins with 584 GO terms. GO terms are classed by the most significant p-value, by this way we retrieved the over-represented pathways expected for our dataset: GO terms relative to hormone activity and glycolytic process (that annotate respectively 33 and 27 proteins of the dataset). We have to note a lack of annotation for 12 proteins of the sample dataset, and a lack of annotation relative to glycolysis for 4 proteins (28 of the 33 expected proteins related to the glycolysis were annotated). This lack of annotations is related to our choice to use only GO terms that have been agreed by review curator in the “Basic Analysis” (no annotation with IEA (Inferred by Electronic Annotation) evidence code). All proteins were annotated when we selected the use of IEA for GO in custom analysis.

ProteINSIDE predicts 85 as potentially secreted. Among the 85 proteins, 81 are confirmed by TargetP and 65 of them are both confirmed by subcellular location and GO terms annotation related to a secretory pathway. By merging shared results from the 3 analysis, we get 78 of the 79 proteins expected on our dataset. ProteINSIDE also predicted 31 proteins as potentially secreted by signal peptide-independent pathways.

The interaction researches between proteins of our sample dataset on 3 databases (BioGrid, IntAct, UniProt) have identified 29 PPi that involved 23 different proteins. As expected, PPi within our dataset linked proteins known to contribute to the pyruvate dehydrogenase complex, the complexes IV and I of the respiratory chain, and also some proteins linked to the glycolysis and the carbohydrate oxidation.

Similar results were obtained from bovine IDs submitted to analyses in bovine species with a custom analysis with settings that used electronic annotation and increased the numbers of PPi databases queried (BioGrid, IntAct, MINT, MatrixDB, STRING, Reactome, InnateDB, I2D and UniProt). This highlights that available data for ruminant species are poorly curated and mainly inferred from electronic annotations. Thus for ruminant IDs, we recommended to proceed to both a custom analysis in the ruminant species and to a basic analysis in a well annotated species (as Rat, Mouse or Human) and to compare the results.

Conclusion

In this work we present the performances of ProteINSIDE, a new powerful workflow which gather tools and public databases to retrieve biological information of genes or proteins lists from 6 species (Bovine, Ovine, Caprine, Human, Rat, and Murine). The presented web service has correctly identified a dataset of 133 proteins, has excluded a duplicate query and has retrieved biological information for each protein. According to our dataset, ProteINSIDE properly annotates the proteins related to the glycolysis, the proteins known as hormones, and the putatively secreted proteins. ProteINSIDE has revealed the most common pathways related to our dataset by creating networks from PPi within the dataset. Each result is easily accessible and downloadable. ProteINSIDE offers a great support to analyse a large quantity of data from genomic and proteomic studies. ProteINSIDE is also the unique web service that makes all of these analyses using ruminant IDs.

Acknowledgements

This work was supported by the regional council of Auvergne in France and APIS-GENE.

References

- Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F.S., Ceol, A., Chautard, E., Dana, J.M., De Las Rivas, J., Dumousseau, M., Galeota, E., Gaulton, A., Goll, J., Hancock, R.E., Isserlin, R., Jimenez, R.C., Kerssemakers, J., Khadake, J., Lynn, D.J., 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, G.D., Cesareni, G., Donaldson, I.M., Eisenberg, D., Kleywegt, G.J., Overington, J., Ricard-Blum, S., Tyers, M., Albrecht, M. and Hermjakob, H., 2011. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat Methods* 8: 528-529.
- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B., 2010. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* 4: 1093-1109.
- Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005-1016.
- Huntley, R.P., Binns, D., Dimmer, E., Barrell, D., O'Donovan, C. and Apweiler, R., 2009. QuickGO: a user tutorial for the web-based Gene Ontology browser. *Database (Oxford)* 2009: bap010.
- Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q. and Bader, G.D., 2010. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26: 2347-2348.
- Nickel, W., 2003. The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *European Journal of Biochemistry* 270: 2109-2119.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785-786.
- Ramirez, F., Lawyer, G. and Albrecht, M., 2012. Novel search method for the discovery of functional relationships. *Bioinformatics* 28: 269-276.



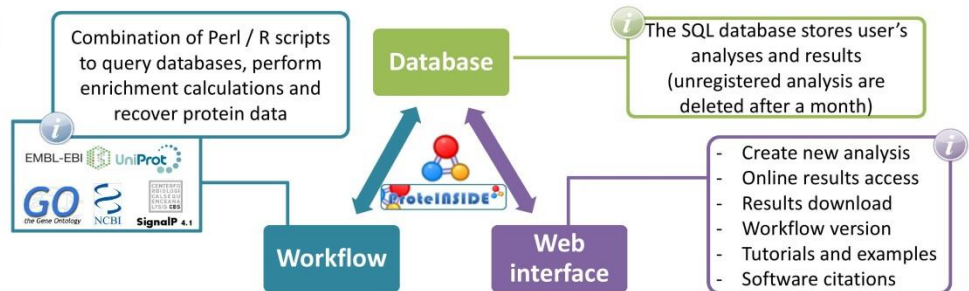
Proteomics data from ruminants easily investigated using ProteINSIDE web service

**KASPRIC Nicolas^{1,2}, REICHSTADT Matthieu^{1,2}, PICARD Brigitte^{1,2},
TOURNAYRE Jérémy^{1,2} and BONNET Muriel^{1,2}**
¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France
² Clermont Université, VetAgro Sup, UMR1213 Herbivores, BP 10448, F-63000, Clermont-Ferrand, France

Given the increasing amount of proteomic data produced, a strategy to perform a systemic and integrative analysis of protein biological meanings is to develop an all-in-one system: ProteINSIDE.

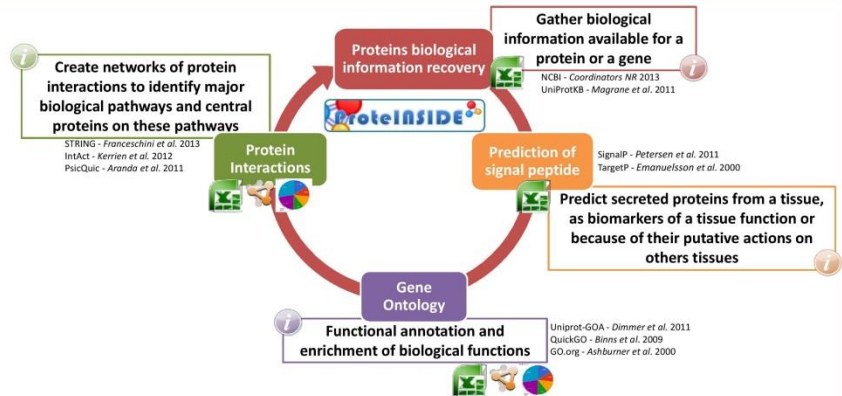
1 Conception

We designed ProteINSIDE as a fast online tool that only requires an internet browser. ProteINSIDE is composed of a database, a web interface, and a workflow.



2 Biological meaning

ProteINSIDE imports accession numbers (usable in other web services) and biological meanings, identifies over- and under-represented biological functions, predicts secreted proteins, and constructs networks from proteins / genes IDs. ProteINSIDE gives a complete overview of a dataset using simple query.



3 How to use it?

- Upload a list of proteins and genes IDs
- Choose a basic (settings locked) or a custom (with your own settings) analysis
- Visualize interactive results

FIC List of proteins/genes



ProteINSIDE worked using 6 species

Farm species:

- Bovine
- Sheep
- Goat



Model species:

- Human
- Mouse
- Rat



ProteINSIDE is an all-in-one tool to automatically query and analyse genomics or proteomics data both in farm and models species. An added-value of ProteINSIDE is the easy and interactive view of the results.



Annexe A.7 : Résumé et présentation orale

Proteome data mining using ProteINSIDE online tool

Nicolas KASPRIC^{1, 2}, Matthieu REICHSTADT^{1, 2}, Brigitte PICARD^{1, 2}, Jérémy TOURNAYRE^{1, 2} and Muriel BONNET^{1, 2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213

Journées Ouvertes en Biologie, Informatique & Mathématiques (JOBIM), Clermont-Ferrand, France, July 6-9 2015 (and oral communication).

Proteome data mining using ProteINSIDE online tool

Nicolas KASPRIC^{1,2}, Matthieu REICHSTADT^{1,2}, Brigitte PICARD^{1,2}, Jérémy TOURNAYRE^{1,2}
and Muriel BONNET^{1,2}

¹ INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

² Clermont Université, VetAgro Sup, UMR1213 Herbivores, F-63000, Clermont-Ferrand, France

Summary

Genomic and proteomic experiments are widely acknowledged to produce a huge amount of data to be analyzed. The challenge for scientist working on ruminant species is to extract meaningful biological context for proteins or genes to generate new hypothesis. This challenge is currently difficult because of the lack of an integrative workflow that hinders the efficiency and the robustness of data mining, thus scientists must use several tools often not designed for ruminant studies. We designed the online tool ProteINSIDE ([Kaspric et al., 2015](#)) that provides an overview of the biological information stored in public databases (NCBI and UniProt) or provided by annotations according to the Gene Ontology. It also predicts proteins that are secreted to search for proteins that mediate signalization between cells or tissues, and analyses protein-protein interactions to identify proteins contributing to a process or to visualize functional pathways. Using lists of proteins or genes as a unique input, ProteINSIDE is an original all-in-one tool to mine genomic and proteomic data from cattle, sheep, goat, human, rat, and mouse. A user-friendly web interface provides access to create analyzes and view results that can be sorted online or directly downloaded as several file formats (image, pdf, tabular, spreadsheet, fasta...). ProteINSIDE's database collects and stores the biological information required to the efficient functioning of ProteINSIDE and is monthly updated to gather the last available knowledge. ProteINSIDE was successfully bench tested with 1000 protein identifiers from each species by comparison with DAVID, BioMyn, and AgBase designed for information retrieval and annotation, as well as with PrediSi, and Phobius that predict secreted proteins. ProteINSIDE is freely available using a simple internet browser at www.proteinside.org (an example of results is provided on the home page).

Résumé

Les expérimentations de génomique et de protéomique produisent de grandes quantités de données à analyser. Le défi pour les scientifiques travaillant sur les ruminants est d'extraire de ces expérimentations une information biologique pertinente afin de générer de nouvelles hypothèses de recherche. Actuellement, l'absence d'outils automatisés dédiés aux ruminants entrave l'exploration efficace et robuste de données disponibles. Les scientifiques utilisent donc différents outils qui ne sont pas forcément conçus pour l'analyse des données issues des ruminants. Nous avons conçu l'outil en ligne ProteINSIDE ([Kaspric et al., 2015](#)) qui synthétise l'information biologique stockée dans les bases de données publiques (NCBI et UniProt) ou fournie par les annotations fonctionnelles issues de l'ontologie des gènes. Il prédit aussi les protéines qui sont sécrétées et qui interviennent dans la signalisation entre les cellules ou tissus, et identifie les interactions entre protéines pour identifier et visualiser celles qui contribuent à un même processus biologique. ProteINSIDE est un outil tout-en-un pour la fouille des données de génomique et de protéomique chez le bovin, l'ovin, le caprin, l'homme, le rat et la souris. L'interface web permet la création d'analyses et la visualisation des résultats en temps réel ou leurs téléchargement dans de nombreux formats (images, pdf, tableaux, données brutes, fasta...). La base de données de ProteINSIDE recueille et stocke l'information biologique nécessaire pour un fonctionnement efficace. Elle est mise à jour mensuellement afin de recueillir les dernières connaissances. ProteINSIDE a été testé avec des jeux de données de 1000 protéines par espèce et a été comparé avec succès à DAVID, BioMyn et AgBase conçus pour la recherche d'information et l'annotation, ainsi qu'à PrediSi et Phobius qui prédisent les protéines sécrétées. ProteINSIDE est disponible gratuitement en utilisant un simple navigateur internet à l'adresse www.proteinside.org (un exemple de résultats est visible sur la page d'accueil).

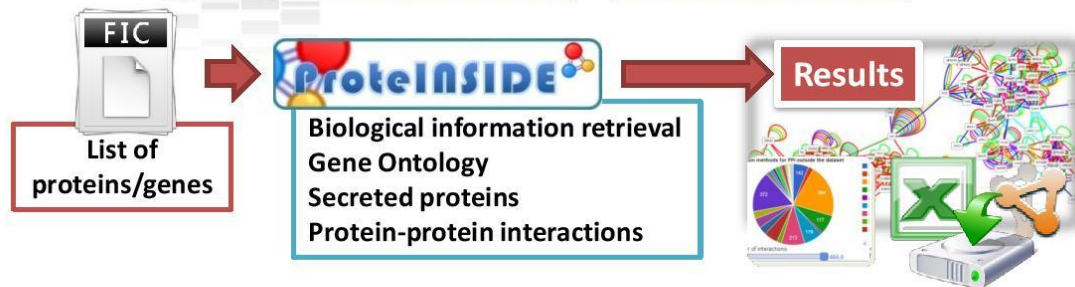
1. Kaspric N, Picard B, Reichstadt M, Tournayre J, Bonnet M: **ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses.** *PLoS One* 2015, **10**:e0128086.

Annexe A.8 : Plaquette de présentation de ProteINSIDE

Proteomics data from ruminants easily investigated using ProteINSIDE web service

Nicolas KASPRIC, Brigitte PICARD, Matthieu REICHSTADT,
Jérémy TOURNAYRE and Muriel BONNET

Contact: nicolas.kaspric@clermont.inra.fr and muriel.bonnet@clermont.inra.fr






ProteINSIDE works using 6 species

Target species:

- Bovine 
- Sheep 
- Goat 

Model species:

- Human 
- Mouse 
- Rat 

ProteINSIDE is an all-in-one tool to automatically query and analyse genomics or proteomics data both in farm and models species. An added-value of ProteINSIDE is the easy and interactive view of the results.

www.proteinside.org

KASPRIC *et al.* 2015 - PlosOne.

ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses.



INRA, UMR1213 Herbivore,
FRANCE
Clermont Université, VetAgro-Sup,
UMR1213 Herbivore, FRANCE



REFERENCES

Références

A

Agapito, G., Guzzi, P.H. and Cannataro, M. Visualization of protein interaction networks: problems and solutions. *BMC Bioinformatics* 2013;14 Suppl 1:S1.

Ailhaud, G., Grimaldi, P. and Negrel, R. Cellular and molecular aspects of adipose tissue development. *Annu. Rev. Nutr.* 1992;12:207-233.

Al-Fakhri, N., Chavakis, T., Schmidt-Woll, T., Huang, B., Cherian, S.M., Bobryshev, Y.V., Lord, R.S., Katz, N. and Preissner, K.T. Induction of apoptosis in vascular cells by plasminogen activator inhibitor-1 and high molecular weight kininogen correlates with their anti-adhesive properties. *Biol. Chem.* 2003;384(3):423-435.

Al-Shahrour, F., Diaz-Uriarte, R. and Dopazo, J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* 2004;20(4):578-580.

Al-Shahrour, F., Minguez, P., Tarraga, J., Medina, I., Alloza, E., Montaner, D. and Dopazo, J. FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res* 2007;35(Web Server issue):W91-96.

Aladag, A.E. and Erten, C. SPINAL: scalable protein interaction network alignment. *Bioinformatics* 2013;29(7):917-924.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-410.

Amthor, H., Huang, R.J., McKinnell, I., Christ, B., Kambadur, R., Sharma, M. and Patel, K. The regulation and action of myostatin as a negative regulator of muscle development during avian embryogenesis. *Dev. Biol.* 2002;251(2):241-257.

Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F.S., Ceol, A., Chautard, E., Dana, J.M., De Las Rivas, J., Dumousseau, M., Galeota, E., Gaulton, A., Goll, J., Hancock, R.E., Isserlin, R., Jimenez, R.C., Kerssemakers, J., Khadake, J., Lynn, D.J., 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, G.D., Cesareni, G., Donaldson, I.M., Eisenberg, D., Kleywegt, G.J., Overington, J., Ricard-Blum, S., Tyers, M., Albrecht, M. and Hermjakob, H. PSICQUIC and PSIScore: accessing and scoring molecular interactions. *Nature methods* 2011;8(7):528-529.

Arita, Y., Kihara, S., Ouchi, N., Maeda, K., Kuriyama, H., Okamoto, Y., Kumada, M., Hotta, K., Nishida, M., Takahashi, M., Nakamura, T., Shimomura, I., Muraguchi, M., Ohmoto, Y., Funahashi, T. and Matsuzawa, Y. Adipocyte-derived plasma protein adiponectin acts as a platelet-derived growth factor-BB-binding protein and regulates growth factor-induced common postreceptor signal in vascular smooth muscle cell. *Circulation* 2002;105(24):2893-2898.

Arita, Y., Kihara, S., Ouchi, N., Maeda, K., Kuriyama, H., Okamoto, Y., Kumada, M., Hotta, K., Nishida, M., Takahashi, M., Nakamura, T., Shimomura, I., Muraguchi, M., Ohmoto, Y., Funahashi, T. and Matsuzawa, Y. Adipocyte-derived plasma protein adiponectin acts as a platelet-derived growth factor-BB-binding protein and regulates growth factor-induced common postreceptor signal in vascular smooth muscle cell. *Circulation* 2002;105(24):2893-2898.

Arnold, H.H. and Winter, B. Muscle differentiation: more complexity to the network of myogenic regulators. *Curr. Opin. Genet. Dev.* 1998;8(5):539-544.

Artaza, J.N., Bhasin, S., Magee, T.R., Reisz-Porszasz, S., Shen, R., Groome, N.P., Fareez, M.M. and Gonzalez-Cadavid, N.F. Myostatin inhibits myogenesis and promotes adipogenesis in C3H 10T(1/2) mesenchymal multipotent cells. *Endocrinology* 2005;146(8):3547-3557.

Artaza, J.N., Bhasin, S., Magee, T.R., Reisz-Porszasz, S., Shen, R.Q., Groome, N.P., Fareez, M.M. and Gonzalez-Cadavid, N.F. Myostatin inhibits myogenesis and promotes adipogenesis in C3H 10T(1/2) mesenchymal multipotent cells. *Endocrinology* 2005;146(8):3547-3557.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. and Stockinger, H. ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res* 2012;40(Web Server issue):W597-603.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 2000;25(1):25-29.

Austin, A.S., Haas, K.M., Naugler, S.M., Bajer, A.A., Garcia-Tapia, D. and Estes, D.M. Identification and characterization of a novel regulatory factor: IgA-inducing protein. *J. Immunol.* 2003;171(3):1336-1342.

B

Bader, G.D. and Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2.

Baerga, R., Zhang, Y., Chen, P.H., Goldman, S. and Jin, S.K. Targeted deletion of autophagy-related 5 (atg5) impairs adipogenesis in a cellular model and in mice. *Autophagy* 2009;5(8):1118-1130.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S. and Soboleva, A. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013;41(Database issue):D991-995.

Baudot, A., Martin, D., Mouren, P., Chevenet, F., Guenoche, A., Jacq, B. and Brun, C. PRODISTIN Web Site: a tool for the functional classification of proteins from interaction networks. *Bioinformatics* 2006;22(2):248-250.

Baudot, A., Souiai, O. and Brun, C. Network analysis and protein function prediction with the PRODISTIN Web site. *Methods in molecular biology* 2012;804:313-326.

Bauer, S., Grossmann, S., Vingron, M. and Robinson, P.N. Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 2008;24(14):1650-1651.

Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A.P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R.L., Yoon, C., Schiess, R., Schmidt, A., Mirkowska, P., Hartlova, A., Van Eyk, J.E., Bourquin, J.P., Aebersold, R., Boheler, K.R., Zandstra, P. and Wollscheid, B. A mass spectrometric-derived cell surface protein atlas. *PLoS one* 2015;10(3):e0121314.

Beissbarth, T. and Speed, T.P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 2004;20(9):1464-1465.

Ben Jemaa, A., Bouraoui, Y., Sallami, S., Banasr, A., Nouira, Y., Horchani, A. and Oueslati, R. Cellular distribution and heterogeneity of Psa and Psma expression in normal, hyperplasia and human prostate cancer. *Tunis. Med.* 2013;91(7):458-463.

Bendtsen, J.D., Jensen, L.J., Blom, N., Von Heijne, G. and Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* 2004;17(4):349-356.

Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 1995;57(1):289-300.

Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. GenBank. *Nucleic Acids Res* 2015;43(Database issue):D30-35.

Berge, C. Théorie des graphes et ses applications. 1958.

Berglund, E.C., Nystedt, B. and Andersson, S.G. Computational resources in infectious disease: limitations and challenges. *PLoS Comput. Biol.* 2009;5(10):e1000481.

- Bianco, P., Sacchetti, B. and Riminucci, M. Osteoprogenitors and the hematopoietic microenvironment. *Best Pract. Res. Clin. Haematol.* 2011;24(1):37-47.
- Billon, N., Iannarelli, P., Monteiro, M.C., Glavieux-Pardanaud, C., Richardson, W.D., Kessar, N., Dani, C. and Dupin, E. The generation of adipocytes by the neural crest. *Development* 2007;134(12):2283-2292.
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z. and Galon, J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25(8):1091-1093.
- Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C. and Apweiler, R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 2009;25(22):3045-3046.
- Biressi, S., Tagliafico, E., Lamorte, G., Monteverde, S., Tenedini, E., Roncaglia, E., Ferrari, S., Ferrari, S., Cusella-De Angelis, M.G., Tajbakhsh, S. and Cossu, G. Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells. *Dev. Biol.* 2007;304(2):633-651.
- Biron, E. and Bedard, F. Recent progress in the development of protein-protein interaction inhibitors targeting androgen receptor-coactivator binding in prostate cancer. *J. Steroid Biochem. Mol. Biol.* 2015.
- Blake, J.A. and Bult, C.J. Beyond the data deluge: data integration and bio-ontologies. *Journal of biomedical informatics* 2006;39(3):314-320.
- Bocquier, F., Guillouet, P., Barillet, F. and Chilliard, Y. Comparison of three methods for the in vivo estimation of body composition in dairy ewes. *Annales De Zootechnie* 1999;48(4):297-308.
- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal* 2010;4(7):1093-1109.
- Bonnet, M., Cassar-Malek, I., Chilliard, Y. and Picard, B. Ontogenesis of muscle and adipose tissues and their interactions in ruminants and other species. *Animal : an international journal of animal bioscience* 2010;4(7):1093-1109.
- Bonnet, M., Faulconnier, Y., Flechet, J., Hocquette, J.F., Leroux, C., Langin, D., Martin, P. and Chilliard, Y. Messenger RNAs encoding lipoprotein lipase, fatty acid synthase and hormone-sensitive lipase in the adipose tissue of underfed-refed ewes and cows. *Reprod. Nutr. Dev.* 1998;38(3):297-307.
- Bonnet, M., Faulconnier, Y., Hocquette, J.F., Bocquier, F., Leroux, C., Martin, P. and Chilliard, Y. Nutritional status induces divergent variations of GLUT4 protein content, but not lipoprotein lipase activity, between adipose tissues and muscles in adult cattle. *The British journal of nutrition* 2004;92(4):617-625.
- Bonnet, M., Louveau, I., Cassar-Malek, I., Lefaucheur, L. and Rescan, P.Y. Comprendre le développement des muscles et des tissus adipeux : un préalable pour maîtriser les qualités des carcasses et des produits des animaux d'élevage. *INRA Productions Animales* 2015;28(2):137-150.
- Boone, C., Mouro, J., Gregoire, F. and Remacle, C. The adipose conversion process: regulation by extracellular and intracellular factors. *Reprod. Nutr. Dev.* 2000;40(4):325-358.
- Bouley, J., Chambon, C. and Picard, B. Mapping of bovine skeletal muscle proteins using two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* 2004;4(6):1811-1824.
- Brambilla, F., Lavatelli, F., Di Silvestre, D., Valentini, V., Palladini, G., Merlini, G. and Mauri, P. Shotgun Protein Profile of Human Adipose Tissue and Its Changes in Relation to Systemic Amyloidosis. *J. Proteome Res.* 2013;12(12):5642-5655.
- Braun, P. and Gingras, A.C. History of protein-protein interactions: from egg-white to complex networks. *Proteomics* 2012;12(10):1478-1498.
- Braun, T., Bober, E., Rudnicki, M.A., Jaenisch, R. and Arnold, H.H. MyoD expression marks the onset of skeletal myogenesis in Myf-5 mutant mice. *Development* 1994;120(11):3083-3092.
- Brazas, M.D., Yim, D.S., Yamada, J.T. and Ouellette, B.F. The 2011 Bioinformatics Links Directory update: more resources, tools and databases and features to empower the bioinformatics community. *Nucleic Acids Res* 2011;39(Web Server issue):W3-7.

Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S. and Lynn, D.J. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res* 2013;41(Database issue):D1228-1233.

Brooksbank, C., Bergman, M.T., Apweiler, R., Birney, E. and Thornton, J. The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Res* 2014;42(Database issue):D18-25.

Brown, K.R. and Jurisica, I. Online predicted human interaction database. *Bioinformatics* 2005;21(9):2076-2082.

Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Royer, N., Vinh, J., Pernet, P., Vaubourdolle, M. and Baudin, B. Proteomics of human umbilical vein endothelial cells applied to etoposide-induced apoptosis. *Proteomics* 2005;5(15):3876-3884.

Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D. and Relaix, F. The formation of skeletal muscle: from somite to limb. *J. Anat.* 2003;202(1):59-68.

Buraschi, S., Neill, T., Goyal, A., Poluzzi, C., Smythies, J., Owens, R.T., Schaefer, L., Torres, A. and Iozzo, R.V. Decorin causes autophagy in endothelial cells via Peg3. *Proc. Natl. Acad. Sci. U. S. A.* 2013;110(28):E2582-E2591.

Buraschi, S., Neill, T., Goyal, A., Poluzzi, C., Smythies, J., Owens, R.T., Schaefer, L., Torres, A. and Iozzo, R.V. Decorin causes autophagy in endothelial cells via Peg3. *Proceedings of the National Academy of Sciences of the United States of America* 2013;110(28):E2582-2591.

Butler-Browne, G. and Bigard, A.X. Muscle ageing and preventive effects of regular exercise. *Sci. Sports* 2006;21(4):184-193.

Byyny, R.L. The data deluge: the information explosion in medicine and science. *The Pharos of Alpha Omega Alpha-Honor Medical Society. Alpha Omega Alpha* 2012;75(2):2-5.

C

Caccia, D., Dugo, M., Callari, M. and Bongarzone, I. Bioinformatics tools for secretome analysis. *Biochim. Biophys. Acta* 2013;1834(11):2442-2453.

Cannata, N., Merelli, E. and Altman, R.B. Time to organize the bioinformatics resourceome. *PLoS Comput. Biol.* 2005;1(7):e76.

Cannon, B. and Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiol. Rev.* 2004;84(1):277-359.

Caprera, A., Lazzari, B., Stella, A., Merelli, I., Caetano, A.R. and Mariani, P. GoSh: a web-based database for goat and sheep EST sequences. *Bioinformatics* 2007;23(8):1043-1045.

Caraux, G. and Pinloche, S. PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2005;21(7):1280-1281.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Ami, G.O.H. and Web Presence Working, G. AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009;25(2):288-289.

Carneiro, D.G., Clarke, T., Davies, C.C. and Bailey, D. Identifying novel protein interactions: proteomic methods, optimisation approaches and data analysis pipelines. *Methods* 2015.

Carpenter, B., MacKay, C., Alnabulsi, A., MacKay, M., Telfer, C., Melvin, W.T. and Murray, G.I. The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression. *Biochim. Biophys. Acta* 2006;1765(2):85-100.

Casteilla, L., Champigny, O., Bouillaud, F., Robelin, J. and Ricquier, D. Sequential changes in the expression of mitochondrial protein mRNA during the development of brown adipose tissue in bovine and ovine species. Sudden occurrence of uncoupling protein mRNA during embryogenesis and its disappearance after birth. *Biochem. J.* 1989;257(3):665-671.

Chandra, D., Choy, G. and Tang, D.G. Cytosolic accumulation of HSP60 during apoptosis with or without apparent mitochondrial release: evidence that its pro-apoptotic or pro-survival functions

involve differential interactions with caspase-3. *The Journal of biological chemistry* 2007;282(43):31289-31301.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K. and Tyers, M. The BioGRID interaction database: 2015 update. *Nucleic Acids Res* 2015;43(Database issue):D470-478.

Chau, Y.Y., Bandiera, R., Serrels, A., Martinez-Estrada, O.M., Qing, W., Lee, M., Slight, J., Thornburn, A., Berry, R., McHaffie, S., Stimson, R.H., Walker, B.R., Chapuli, R.M., Schedl, A. and Hastie, N. Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. *Nat. Cell Biol.* 2014;16(4):367-+.

Chaze, T.: Université Blaise Pascal; 2007. Evolution du protéome musculaire au cours de la myogenèse chez le bovin.

Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B. In vivo proteome dynamics during early bovine myogenesis. *Proteomics* 2008;8(20):4236-4248.

Chaze, T., Meunier, B., Chambon, C., Jurie, C. and Picard, B. Proteome dynamics during contractile and metabolic differentiation of bovine foetal muscle. *Animal : an international journal of animal bioscience* 2009;3(7):980-1000.

Chen, J., Bardes, E.E., Aronow, B.J. and Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009;37(Web Server issue):W305-311.

Chen, Y., Zhang, Y., Yin, Y., Gao, G., Li, S., Jiang, Y., Gu, X. and Luo, J. SPD--a web-based secreted protein database. *Nucleic Acids Res* 2005;33(Database issue):D169-173.

Chen, Y.P. and Chen, F. Identifying targets for drug discovery using bioinformatics. *Expert Opin. Ther. Targets* 2008;12(4):383-389.

Chilliard, Y. Dietary fat and adipose tissue metabolism in ruminants, pigs, and rodents: a review. *J. Dairy Sci.* 1993;76(12):3897-3931.

Chilliard, Y., Bonnet, M., Delavaud, C., Faulconnier, Y., Leroux, C., Djiane, J. and Bocquier, F. Leptin in ruminants. Gene expression in adipose tissue and mammary gland, and regulation of plasma concentration. *Domest. Anim. Endocrinol.* 2001;21(4):271-295.

Chilliard, Y., Delavaud, C. and Bonnet, M. Leptin expression in ruminants: nutritional and physiological regulations in relation with energy metabolism. *Domest. Anim. Endocrinol.* 2005;29(1):3-22.

Chilliard, Y., Ferlay, A., Faulconnier, Y., Bonnet, M., Rouel, J. and Bocquier, F. Adipose tissue metabolism and its role in adaptations to undernutrition in ruminants. *Proc. Nutr. Soc.* 2000;59(1):127-134.

Chokkalingam, V., Tel, J., Wimmers, F., Liu, X., Semenov, S., Thiele, J., Figdor, C.G. and Huck, W.T. Probing cellular heterogeneity in cytokine-secreting immune cells using droplet-based microfluidics. *Lab on a chip* 2013;13(24):4740-4744.

Chou, K.C. and Shen, H.B. Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat Protoc* 2008;3(2):153-162.

Christ, B. and Brand-Saberi, B. Limb muscle development. *Int. J. Dev. Biol.* 2002;46(7):905-914.

Cianzio, D.S., Topel, D.G., Whitehurst, G.B., Beitz, D.C. and Self, H.L. Adipose tissue growth and cellularity: changes in bovine adipocyte size and number. *Journal of animal science* 1985;60(4):970-976.

Cinti, S. The adipose organ. *Prostaglandins Leukot. Essent. Fatty Acids* 2005;73(1):9-15.

Cinti, S., Frederich, R.C., Zingaretti, M.C., DeMatteis, R., Flier, J.S. and Lowell, B.B. Immunohistochemical localization of leptin and uncoupling protein in white and brown adipose tissue. *Endocrinology* 1997;138(2):797-804.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T. and Bader,

G.D. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2007;2(10):2366-2382.

Coordinators, N.R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2015;43(Database issue):D6-17.

Corton, M., Botella-Carretero, J.I., Lopez, J.A., Camafeita, E., San Millan, J.L., Escobar-Morreale, H.F. and Peral, B. Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. *Human reproduction* 2008;23(3):651-661.

Cote, R., Reisinger, F., Martens, L., Barsnes, H., Vizcaino, J.A. and Hermjakob, H. The Ontology Lookup Service: bigger and better. *Nucleic Acids Res* 2010;38(Web Server issue):W155-160.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L. and D'Eustachio, P. The Reactome pathway knowledgebase. *Nucleic Acids Res* 2014;42(Database issue):D472-477.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Kahari, A.K., Keenan, S., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Aken, B.L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S.M., Spudich, G., Trevanion, S.J., Yates, A., Zerbino, D.R. and Flicek, P. Ensembl 2015. *Nucleic Acids Res* 2015;43(Database issue):D662-669.

D

Dai, Y., Ngo, D., Jacob, J., Forman, L.W. and Faller, D.V. Prohibitin and the SWI/SNF ATPase subunit BRG1 are required for effective androgen antagonist-mediated transcriptional repression of androgen receptor-regulated genes. *Carcinogenesis* 2008;29(9):1725-1733.

Dameron, O., Bettembourg, C. and Le Meur, N. Measuring the evolution of ontology complexity: the gene ontology case study. *PLoS one* 2013;8(10):e75993.

Dani, C. Activins in adipogenesis and obesity. *International journal of obesity* 2013;37(2):163-166.

del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L. and Hermjakob, H. A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res* 2013;41(Web Server issue):W601-606.

DeLany, J.P., Floyd, Z.E., Zvonic, S., Smith, A., Gravois, A., Reiners, E., Wu, X., Kilroy, G., Lefevre, M. and Gimble, J.M. Proteomic analysis of primary cultures of human adipose-derived stem cells: modulation by Adipogenesis. *Mol. Cell. Proteomics* 2005;4(6):731-740.

Delavaud, C., Bocquier, F., Chilliard, Y., Keisler, D.H., Gertler, A. and Kann, G. Plasma leptin determination in ruminants: effect of nutritional status and body fatness on plasma leptin concentration assessed by a specific RIA in sheep. *J. Endocrinol.* 2000;165(2):519-526.

Delavaud, C., Ferlay, A., Faulconnier, Y., Bocquier, F., Kann, G. and Chilliard, Y. Plasma leptin concentration in adult cattle: Effects of breed, adiposity, feeding level, and meal intake. *J. Anim. Sci.* 2002;80(5):1317-1328.

Deutsch, E.W. The PeptideAtlas Project. *Methods in molecular biology* 2010;604:285-296.

Deveaux, V., Cassar-Malek, I. and Picard, B. Comparison of contractile characteristics of muscle from Holstein and double-musled Belgian Blue fetuses. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 2001;131(1):21-29.

Diaz-Ramos, A., Roig-Borrellas, A., Garcia-Melero, A. and Lopez-Aleman, R. alpha-Enolase, a multifunctional protein: its role on pathophysiological situations. *J. Biomed. Biotechnol.* 2012;2012:156795.

- Digby, J.E., Crowley, V.E.F., Sewter, C.P., Whitehead, J.P., Prins, J.B. and O'Rahilly, S. Depot-related and thiazolidinedione-responsive expression of uncoupling protein 2 (UCP2) in human adipocytes. *Int. J. Obes.* 2000;24(5):585-592.
- Dijkstra, E.W. A short introduction to the art of programming. 1971.
- Ding, S.T., McNeel, R.L. and Mersmann, H.J. Expression of porcine adipocyte transcripts: tissue distribution and differentiation in vitro and in vivo. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 1999;123(3):307-318.
- Draeger, A., Weeds, A.G. and Fitzsimons, R.B. Primary, secondary and tertiary myotubes in developing skeletal muscle: a new approach to the analysis of human myogenesis. *J. Neurol. Sci.* 1987;81(1):19-43.
- Du, M., Tong, J., Zhao, J., Underwood, K.R., Zhu, M., Ford, S.P. and Nathanielsz, P.W. Fetal programming of skeletal muscle development in ruminant animals. *J. Anim. Sci.* 2010;88:E51-E60.
- Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 2010;38(Web Server issue):W64-70.
- Duris, M.P., Renand, G. and Picard, B. Genetic variability of foetal bovine myoblasts in primary culture. *Histochem. J.* 1999;31(12):753-760.
- Dutkowski, J., Kramer, M., Surma, M.A., Balakrishnan, R., Cherry, J.M., Krogan, N.J. and Ideker, T. A gene ontology inferred from molecular networks. *Nature biotechnology* 2013;31(1):38-45.
- Dyck, D.J. Adipokines as regulators of muscle metabolism and insulin sensitivity. *Appl. Physiol. Nutr. Metab.* 2009;34(3):396-402.

E-F

- Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 2009;10:48.
- Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007;2(4):953-971.
- Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 2000;300(4):1005-1016.
- Fernandez-Suarez, X.M. and Galperin, M.Y. The 2013 Nucleic Acids Research Database Issue and the online molecular biology database collection. *Nucleic Acids Res* 2013;41(Database issue):D1-7.
- Feve, B. Adipogenesis: cellular and molecular aspects. *Best Pract. Res. Clin. Endocrinol. Metab.* 2005;19(4):483-499.
- Field, D., Tiwari, B., Booth, T., Houten, S., Swan, D., Bertrand, N. and Thurston, M. Open software for biologists: from famine to feast. *Nature biotechnology* 2006;24(7):801-803.
- Fink, J.L., Aturaliya, R.N., Davis, M.J., Zhang, F.S., Hanson, K., Teasdale, M.S., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. and Teasdale, R.D. LOCATE: a mouse protein subcellular localization database. *Nucleic Acids Res.* 2006;34:D213-D217.
- Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C.G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kahari, A.K., Keenan, S., Kulesha, E., Martin, F.J., Maurel, T., McLaren, W.M., Murphy, D.N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S.J., Vullo, A., Wilder, S.P., Wilson, M., Zadissa, A., Aken, B.L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T.J., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D.R. and Searle, S.M. Ensembl 2014. *Nucleic Acids Res* 2014;42(Database issue):D749-755.
- Fliers, E., Kreier, F., Voshol, P.J., Havekes, L.M., Sauerwein, H.P., Kalsbeek, A., Buijs, R.M. and Romijn, J.A. White adipose tissue: getting nervous. *Journal of neuroendocrinology* 2003;15(11):1005-1010.

Florini, J.R., Ewton, D.Z. and Roof, S.L. Insulin-Like Growth Factor-I Stimulates Terminal Myogenic Differentiation by Induction of Myogenin Gene-Expression. *Mol. Endocrinol.* 1991;5(5):718-724.

Forrester, S.J., Kawata, K., Lee, H., Kim, J.S., Sebзда, K., Butler, T., Yingling, V.R. and Park, J.Y. Bioinformatic identification of connective tissue growth factor as an osteogenic protein within skeletal muscle. *Physiological reports* 2014;2(12).

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. and Jensen, L.J. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013;41(Database issue):D808-815.

G

Gagaoua, M., Hafid, K., Boudida, Y., Becila, S., Ouali, A., Picard, B., Boudjellal, A. and Sentandreu, M.A. Caspases and Thrombin Activity Regulation by Specific Serpin Inhibitors in Bovine Skeletal Muscle. *Appl. Biochem. Biotechnol.* 2015;177(2):279-303.

Gagniere, H., Picard, B., Jurie, C. and Geay, Y. Comparative study of metabolic differentiation of foetal muscle in normal and double-muscled cattle. *Meat Sci.* 1997;45(2):145-152.

Gagniere, H., Picard, B., Jurie, C. and Geay, Y. Comparison of foetal metabolic differentiation in three cattle muscles. *Reprod. Nutr. Dev.* 1999;39(1):105-112.

Galperin, M.Y. and Fernandez-Suarez, X.M. The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res* 2012;40(Database issue):D1-8.

Galperin, M.Y., Rigden, D.J. and Fernandez-Suarez, X.M. The 2015 Nucleic Acids Research Database Issue and molecular biology database collection. *Nucleic Acids Res* 2015;43(Database issue):D1-5.

Garcia-Garcia, J., Schleker, S., Klein-Seetharaman, J. and Oliva, B. BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference. *Nucleic Acids Res* 2012;40(Web Server issue):W147-151.

Garin-Shkolnik, T., Rudich, A., Hotamisligil, G.S. and Rubinstein, M. FABP4 Attenuates PPAR gamma and Adipogenesis and Is Inversely Correlated With PPARg in Adipose Tissues. *Diabetes* 2014;63(3):900-911.

Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015;43(Database issue):D1049-1056.

Ghosh, A.K., Steele, R. and Ray, R.B. Functional domains of c-myc promoter binding protein 1 involved in transcriptional repression and cell growth regulation. *Mol. Cell. Biol.* 1999;19(4):2880-2886.

Giralt, M., Casteilla, L., Vinas, O., Mampel, T., Iglesias, R., Robelin, J. and Villarroya, F. Iodothyronine 5'-Deiodinase Activity as an Early Event of Prenatal Brown-Fat Differentiation in Bovine Development. *Biochem. J.* 1989;259(2):555-559.

Gobeill, J., Pasche, E., Vishnyakova, D. and Ruch, P. Managing the data deluge: data-driven GO category assignment improves while complexity of functional annotation increases. *Database-Oxford* 2013.

Goecks, J., Nekrutenko, A., Taylor, J. and Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology* 2010;11(8):R86.

Goldfarb, L.G., Park, K.Y., Cervenakova, L., Gorokhova, S., Lee, H.S., Vasconcelos, O., Nagle, J.W., Semino-Mora, C., Sivakumar, K. and Dalakas, M.C. Missense mutations in desmin associated with familial cardiac and skeletal myopathy. *Nature genetics* 1998;19(4):402-403.

Goll, J., Rajagopala, S.V., Shiau, S.C., Wu, H., Lamb, B.T. and Uetz, P. MPIDB: the microbial protein interaction database. *Bioinformatics* 2008;24(15):1743-1744.

- Goncalves, E. and Saez-Rodriguez, J. Cyrface: An interface from Cytoscape to R that provides a user interface to R packages. *F1000Research* 2013;2:192.
- Gonzalez, J.M., Camacho, L.E., Ebarb, S.M., Swanson, K.C., Vonnahme, K.A., Stelzleni, A.M. and Johnson, S.E. Realimentation of nutrient restricted pregnant beef cows supports compensatory fetal muscle growth. *Journal of animal science* 2013;91(10):4797-4806.
- Gotoh, T., Albrecht, E., Teuscher, F., Kawabata, K., Sakashita, K., Iwamoto, H. and Wegner, J. Differences in muscle and fat accretion in Japanese Black and European cattle. *Meat Sci.* 2009;82(3):300-308.
- Gotoh, T., Albrecht, E., Teuscher, F., Kawabata, K., Sakashita, K., Iwamoto, H. and Wegner, J. Differences in muscle and fat accretion in Japanese Black and European cattle. *Meat science* 2009;82(3):300-308.
- Gregoire, F.M. Adipocyte differentiation: from fibroblast to endocrine cell. *Exp. Biol. Med. (Maywood)* 2001;226(11):997-1002.
- Gregoire, F.M., Smas, C.M. and Sul, H.S. Understanding adipocyte differentiation. *Physiol. Rev.* 1998;78(3):783-809.
- Gregoire, F.M., Smas, C.M. and Sul, H.S. Understanding adipocyte differentiation. *Physiol. Rev.* 1998;78(3):783-809.
- Guo, W., Flanagan, J., Jasuja, R., Kirkland, J., Jiang, L. and Bhasin, S. The effects of myostatin on adipogenic differentiation of human bone marrow-derived mesenchymal stem cells are mediated through cross-communication between Smad3 and Wnt/beta-catenin signaling pathways. *The Journal of biological chemistry* 2008;283(14):9136-9145.
- Gustafson, B. and Smith, U. Cytokines promote Wnt signaling and inflammation and impair the normal differentiation and lipid accumulation in 3T3-L1 preadipocytes. *The Journal of biological chemistry* 2006;281(14):9507-9516.

H

- Hanson, A.D., Pribat, A., Waller, J.C. and de Crecy-Lagard, V. 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list--and how to find it. *Biochem. J.* 2010;425(1):1-11.
- Haoudi, A. and Bensmail, H. Bioinformatics and data mining in proteomics. *Expert Rev. Proteomics* 2006;3(3):333-343.
- Hasty, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N. and Klein, W.H. Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature* 1993;364(6437):501-506.
- Hausman, D.B., DiGirolamo, M., Bartness, T.J., Hausman, G.J. and Martin, R.J. The biology of white adipocyte proliferation. *Obes. Rev.* 2001;2(4):239-254.
- He, L.W., Vichev, K., Macharia, R., Huang, R.J., Christ, B., Patel, K. and Amthor, H. Activin A inhibits formation of skeletal muscle during chick development. *Anat. Embryol.* 2005;209(5):401-407.
- Henry, V.J., Bandrowski, A.E., Pepin, A.S., Gonzalez, B.J. and Desfeux, A. OMICtools: an informative directory for multi-omic data analysis. *Database : the journal of biological databases and curation* 2014;2014.
- Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C., Roechert, B., Poux, S., Jung, E., Mersch, H., Kersey, P., Lappe, M., Li, Y., Zeng, R., Rana, D., Nikolski, M., Husi, H., Brun, C., Shanker, K., Grant, S.G., Sander, C., Bork, P., Zhu, W., Pandey, A., Brazma, A., Jacq, B., Vidal, M., Sherman, D., Legrain, P., Cesareni, G., Xenarios, I., Eisenberg, D., Steipe, B., Hogue, C. and Apweiler, R. The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. *Nature biotechnology* 2004;22(2):177-183.
- Hiller, K., Grote, A., Scheer, M., Munch, R. and Jahn, D. PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* 2004;32(Web Server issue):W375-379.

- Hirai, S., Matsumoto, H., Hino, N., Kawachi, H., Matsui, T. and Yano, H. Myostatin Inhibits Differentiation of Bovine Preadipocyte. *Domestic Animal Endocrinology* 2007;32(1).
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O. and Rhee, S.Y. Big data: The future of biocuration. *Nature* 2008;455(7209):47-50.
- Hu, Z.L., Park, C.A., Wu, X.L. and Reecy, J.M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 2013;41(D1):D871-D879.
- Huang, D.W., Sherman, B.T. and Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4(1):44-57.
- Huang, W.L. Ranking Gene Ontology terms for predicting non-classical secretory proteins in eukaryotes and prokaryotes. *Journal of theoretical biology* 2012;312:105-113.
- Hudson, N.J., Lyons, R.E., Reverter, A., Greenwood, P.L. and Dalrymple, B.P. Inferring the in vivo cellular program of developing bovine skeletal muscle from expression data. *Gene expression patterns : GEP* 2013;13(3-4):109-125.
- Huntley, R.P., Binns, D., Dimmer, E., Barrell, D., O'Donovan, C. and Apweiler, R. QuickGO: a user tutorial for the web-based Gene Ontology browser. *Database : the journal of biological databases and curation* 2009;2009:bap010.
- Huntley, R.P., Sawford, T., Martin, M.J. and O'Donovan, C. Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt. *GigaScience* 2014;3(1):4.
- Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* 2015;43(Database issue):D1057-1063.
- Hwang, S., Son, S.W., Kim, S.C., Kim, Y.J., Jeong, H. and Lee, D. A protein interaction network associated with asthma. *Journal of theoretical biology* 2008;252(4):722-731.

I-J

- Iruelaarispe, M.L., Liska, D.J., Sage, E.H. and Bornstein, P. Differential Expression of Thrombospondin-1, Thrombospondin-2, and Thrombospondin-3 during Murine Development. *Dev. Dyn.* 1993;197(1):40-56.
- Isserlin, R., El-Badrawi, R.A. and Bader, G.D. The Biomolecular Interaction Network Database in PSI-MI 2.5. *Database : the journal of biological databases and curation* 2011;2011:baq037.
- Itoh, N. and Ohta, H. Fgf10: a paracrine-signaling molecule in development, disease, and regenerative medicine. *Curr. Mol. Med.* 2014;14(4):504-509.
- Jin, L., Wang, W. and Fang, G. Targeting protein-protein interaction by small molecules. *Annu. Rev. Pharmacol. Toxicol.* 2014;54:435-456.
- Joo, J.I., Kim, D.H., Choi, J.W. and Yun, J.W. Proteomic analysis for antiobesity potential of capsaicin on white adipose tissue in rats fed with a high fat diet. *J. Proteome Res.* 2010;9(6):2977-2987.

K

- Kalathur, R.K., Pinto, J.P., Hernandez-Prieto, M.A., Machado, R.S., Almeida, D., Chaurasia, G. and Futschik, M.E. UniHI 7: an enhanced database for retrieval and interactive analysis of human molecular interaction networks. *Nucleic Acids Res* 2014;42(Database issue):D408-414.
- Kall, L., Krogh, A. and Sonnhammer, E.L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 2007;35(Web Server issue):W429-432.
- Kammoun, M., Cassar-Malek, I., Meunier, B. and Picard, B. A simplified immunohistochemical classification of skeletal muscle fibres in mouse. *Eur. J. Histochem.* 2014;58(2):2254.

- Kamradt, M.C., Chen, F., Sam, S. and Cryns, V.L. The small heat shock protein alpha B-crystallin negatively regulates apoptosis during myogenic differentiation by inhibiting caspase-3 activation. *J. Biol. Chem.* 2002;277(41):38731-38736.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 2014;42(Database issue):D199-205.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2015.
- Kanno, Y., Kaneiwa, A., Minamida, M., Kanno, M., Tomogane, K., Takeuchi, K., Okada, K., Ueshima, S., Matsuo, O. and Matsuno, H. The absence of uPAR is associated with the progression of dermal fibrosis. *J. Invest. Dermatol.* 2008;128(12):2792-2797.
- Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., Malone, J., Rustici, G., Williams, E., Parkinson, H. and Brazma, A. Gene expression atlas at the European bioinformatics institute. *Nucleic Acids Res* 2010;38(Database issue):D690-698.
- Kaspric, N., Picard, B., Reichstadt, M., Tournayre, J. and Bonnet, M. Protein function easily investigated by genomics data mining using the ProteINSIDE web service. *Proceedings Iwbbio 2014: International Work-Conference on Bioinformatics and Biomedical Engineering, Vols 1 and 2* 2014:890-U2803.
- Kaspric, N., Picard, B., Reichstadt, M., Tournayre, J. and Bonnet, M. ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. *PLoS one* 2015;10(5):e0128086.
- Kaspric, N., Reichstadt, M., Picard, B., Tournayre, J. and Bonnet, M. Protein Function Easily Investigated by Genomics Data Mining Using the ProteINSIDE Online Tool. *Genomics and Computational Biology* 2015;1(1):e16.
- Kassar-Duchossoy, L., Gayraud-Morel, B., Gomes, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S. Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature* 2004;431(7007):466-471.
- Kawaguchi, N., Nakao, R., Yamaguchi, M., Ogawa, D. and Matsuoka, R. TGF-beta superfamily regulates a switch that mediates differentiation either into adipocytes or myocytes in left atrium derived pluripotent cells (LA-PCS). *Biochem. Biophys. Res. Commun.* 2010;396(3):619-625.
- Kawwass, J.F., Summer, R. and Kallen, C.B. Direct effects of leptin and adiponectin on peripheral reproductive tissues: a critical review. *Mol. Hum. Reprod.* 2015.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadrana, S., Chaerkady, R. and Pandey, A. Human Protein Reference Database--2009 update. *Nucleic Acids Res* 2009;37(Database issue):D767-772.
- Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V.H., Nicholas, F.W., Raadsma, H. and International Sheep Genomics, C. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS one* 2009;4(3):e4668.
- Kishioka, Y., Thomas, M., Wakamatsu, J.I., Hattori, A., Sharma, M., Kambadur, R. and Nishimura, T. Decorin enhances the proliferation and differentiation of myogenic cells through suppressing myostatin activity. *J. Cell. Physiol.* 2008;215(3):856-867.
- Klee, E.W., Carlson, D.F., Fahrenkrug, S.C., Ekker, S.C. and Ellis, L.B. Identifying secretomes in people, pufferfish and pigs. *Nucleic Acids Res* 2004;32(4):1414-1421.
- Klee, E.W. and Ellis, L.B. Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* 2005;6:256.
- Klingstrom, T. and Plewczynski, D. Protein-protein interaction and pathway databases, a graphical review. *Brief Bioinform* 2011;12(6):702-713.

Koenig, T., Menze, B.H., Kirchner, M., Monigatti, F., Parker, K.C., Patterson, T., Steen, J.J., Hamprecht, F.A. and Steen, H. Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J. Proteome Res.* 2008;7(9):3708-3717.

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., Megy, K., Pilicheva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U. and Brazma, A. ArrayExpress update-simplifying data submissions. *Nucleic Acids Res.* 2015;43(D1):D1113-D1116.

Komolka, K., Albrecht, E., Schering, L., Brenmoehl, J., Hoeflich, A. and Maak, S. Locus characterization and gene expression of bovine FNDC5: is the myokine irisin relevant in cattle? *PLoS one* 2014;9(1):e88060.

Kopp, A., Bala, M., Buechler, C., Falk, W., Gross, P., Neumeier, M., Scholmerich, J. and Schaffler, A. C1q/TNF-related protein-3 represents a novel and endogenous lipopolysaccharide antagonist of the adipose tissue. *Endocrinology* 2010;151(11):5267-5278.

Kratchmarova, I., Kalume, D.E., Blagoev, B., Scherer, P.E., Podtelejnikov, A.V., Molina, H., Bickel, P.E., Andersen, J.S., Fernandez, M.M., Bunkenborg, J., Roepstorff, P., Kristiansen, K., Lodish, H.F., Mann, M. and Pandey, A. A proteomic approach for identification of secreted proteins during the differentiation of 3T3-L1 preadipocytes to adipocytes. *Mol. Cell. Proteomics* 2002;1(3):213-222.

Kriventseva, E.V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simao, F.A., Pozdnyakov, I.A., Ioannidis, P. and Zdobnov, E.M. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res* 2015;43(Database issue):D250-256.

Kumar, S. and Dudley, J. Bioinformatics software for biologists in the genomics era. *Bioinformatics* 2007;23(14):1713-1717.

Kurys, G., Ambroziak, W. and Pietruszko, R. Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low Km for gamma-aminobutyraldehyde. *The Journal of biological chemistry* 1989;264(8):4715-4721.

L

Landis, M.D., Carstens, G.E., McPhail, E.G., Randel, R.D., Green, K.K., Slay, L. and Smith, S.B. Ontogenic development of brown adipose tissue in Angus and Brahman fetal calves. *Journal of animal science* 2002;80(3):591-601.

Laukens, K., Naulaerts, S. and Berghe, W.V. Bioinformatics approaches for the functional interpretation of protein lists: from ontology term enrichment to network analysis. *Proteomics* 2015;15(5-6):981-996.

Launay, G., Salza, R., Multedo, D., Thierry-Mieg, N. and Ricard-Blum, S. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res* 2015;43(Database issue):D321-327.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 2013;9(8):e1003118.

Le Bihan, M.C., Bigot, A., Jensen, S.S., Dennis, J.L., Rogowska-Wrzęsinska, A., Laine, J., Gache, V., Furling, D., Jensen, O.N., Voit, T., Mouly, V., Coulton, G.R. and Butler-Browne, G. In-depth analysis of the secretome identifies three major independent secretory pathways in differentiating human myoblasts. *Journal of proteomics* 2012;77:344-356.

Lefaucheur, L. A second look into fibre typing--relation to meat quality. *Meat science* 2010;84(2):257-270.

Lefaucheur, L., Edom, F., Ecolan, P. and Butler-Browne, G.S. Pattern of muscle fiber type formation in the pig. *Dev. Dyn.* 1995;203(1):27-41.

Lefterova, M.I. and Lazar, M.A. New developments in adipogenesis. *Trends Endocrinol. Metab.* 2009;20(3):107-114.

- Lehnert, S.A., Reverter, A., Byrne, K.A., Wang, Y., Natrass, G.S., Hudson, N.J. and Greenwood, P.L. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev. Biol.* 2007;7.
- Lemor, A., Mielenz, M., Altmann, M., von Borell, E. and Sauerwein, H. mRNA abundance of adiponectin and its receptors, leptin and visfatin and of G-protein coupled receptor 41 in five different fat depots from sheep. *J Anim Physiol Anim Nutr (Berl)* 2010;94(5):e96-101.
- Letovsky, S. and Kasif, S. Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics* 2003;19:i197-i204.
- Levi, B. and Longaker, M.T. Concise review: adipose-derived stromal cells for skeletal regenerative medicine. *Stem Cells* 2011;29(4):576-582.
- Li, Q.Y., Chen, L., Yan, M.M., Shi, X.J. and Zhong, M.K. Tectorigenin regulates adipogenic differentiation and adipocytokines secretion via PPARgamma and IKK/NF-kappaB signaling. *Pharm. Biol.* 2015:1-9.
- Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardoza, A.P., Santonico, E., Castagnoli, L. and Cesareni, G. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;40(Database issue):D857-861.
- Lievens, S., Eyckerman, S., Lemmens, I. and Tavernier, J. Large-scale protein interactome mapping: strategies and opportunities. *Expert Rev. Proteomics* 2010;7(5):679-690.
- Lilla, J., Stickens, D. and Werb, Z. Metalloproteases and adipogenesis: a weighty subject. *Am. J. Pathol.* 2002;160(5):1551-1554.
- Lim, D., Kim, N.K., Park, H.S., Lee, S.H., Cho, Y.M., Oh, S.J., Kim, T.H. and Kim, H. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *International journal of biological sciences* 2011;7(7):992-1002.
- Lin, K.Y., Wang, H.H., Lai, S.T., Pan, J.P. and Chiang, A.N. beta(2)-glycoprotein I protects J774A.1 macrophages and human coronary artery smooth muscle cells against apoptosis. *J. Cell. Biochem.* 2005;94(3):485-496.
- Lisacek, F., Cohen-Boulakia, S. and Appel, R.D. Proteome informatics II: bioinformatics for comparative proteomics. *Proteomics* 2006;6(20):5445-5466.
- Liu, J.J. and Farmer, S.R. Regulating the balance between peroxisome proliferator-activated receptor gamma and beta-catenin signaling during adipogenesis - A glycogen synthase kinase 3 beta phosphorylation-defective mutant of beta-catenin inhibits expression of a subset of adipogenic genes. *J. Biol. Chem.* 2004;279(43):45020-45027.
- Liu, W.Y., Liu, Y.Q., Lai, X.S. and Kuang, S.H. Intramuscular adipose is derived from a non-Pax3 lineage and required for efficient regeneration of skeletal muscles. *Dev. Biol.* 2012;361(1):27-38.
- Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q. and Bader, G.D. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 2010;26(18):2347-2348.
- Luscombe, N.M., Greenbaum, D. and Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* 2001;40(4):346-358.

M

- Maere, S., Heymans, K. and Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005;21(16):3448-3449.
- Martha, V.S., Liu, Z., Guo, L., Su, Z., Ye, Y., Fang, H., Ding, D., Tong, W. and Xu, X. Constructing a robust protein-protein interaction network by integrating multiple public databases. *BMC Bioinformatics* 2011;12 Suppl 10:S7.
- McCarthy, F.M., Gresham, C.R., Buza, T.J., Chouvarine, P., Pillai, L.R., Kumar, R., Ozkan, S., Wang, H., Manda, P., Arick, T., Bridges, S.M. and Burgess, S.C. AgBase: supporting functional modeling in agricultural organisms. *Nucleic Acids Res* 2011;39(Database issue):D497-506.

- McMillan, E.M. and Quadrilatero, J. Autophagy is required and protects against apoptosis during myoblast differentiation. *Biochem. J.* 2014;462(2):267-277.
- Medina, I., Carbonell, J., Pulido, L., Madeira, S.C., Goetz, S., Conesa, A., Tarraga, J., Pascual-Montano, A., Nogales-Cadenas, R., Santoyo, J., Garcia, F., Marba, M., Montaner, D. and Dopazo, J. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res* 2010;38(Web Server issue):W210-213.
- Meinken, J. and Min, J. Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report. *Computational Molecular Biology* 2012;2(1).
- Meinken, J., Walker, G., Cooper, C.R. and Min, X.J. MetazSecKB: the human and animal secretome and subcellular proteome knowledgebase. *Database : the journal of biological databases and curation* 2015;2015.
- Meunier, B., Dumas, E., Piec, I., Bechet, D., Hebraud, M. and Hocquette, J.F. Assessment of hierarchical clustering methodologies for proteomic data mining. *J. Proteome Res.* 2007;6(1):358-366.
- Mhawech, P. 14-3-3 proteins--an update. *Cell Res.* 2005;15(4):228-236.
- Mi, H., Muruganujan, A., Casagrande, J.T. and Thomas, P.D. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 2013;8(8):1551-1566.
- Miegueu, P., Cianflone, K., Richard, D. and St-Pierre, D.H. Motilin stimulates preadipocyte proliferation and differentiation and adipocyte lipid storage. *Am. J. Physiol. Endocrinol. Metab.* 2011;301(5):E758-766.
- Morino, K., Petersen, K.F. and Shulman, G.I. Molecular mechanisms of insulin resistance in humans and their potential links with mitochondrial dysfunction. *Diabetes* 2006;55:S9-S15.
- Muneta, Y., Kikuma, R., Yoshihara, K. and Mori, Y. Cloning, expression, and tissue distribution of bovine interleukin-21. *Vet. Immunol. Immunopathol.* 2003;95(1-2):73-80.
- Munoz-Marmol, A.M., Strasser, G., Isamat, M., Coulombe, P.A., Yang, Y., Roca, X., Vela, E., Mate, J.L., Coll, J., Fernandez-Figueras, M.T., Navas-Palacios, J.J., Ariza, A. and Fuchs, E. A dysfunctional desmin mutation in a patient with severe generalized myopathy. *Proceedings of the National Academy of Sciences of the United States of America* 1998;95(19):11312-11317.
- Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G., 3rd and Finley, R.L., Jr. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res* 2011;39(Database issue):D736-743.
- Murdolo, G., Piroddi, M., Luchetti, F., Tortoioli, C., Canonico, B., Zerbinati, C., Galli, F. and Iuliano, L. Oxidative stress and lipid peroxidation by-products at the crossroad between adipose organ dysregulation and obesity-linked insulin resistance. *Biochimie* 2013;95(3):585-594.
- Muroya, S., Nakajima, I. and Chikuni, K. Amino acid sequences of multiple fast and slow troponin T isoforms expressed in adult bovine skeletal muscles. *Journal of animal science* 2003;81(5):1185-1192.

N

- Nedergaard, J. and Cannon, B. How brown is brown fat? It depends where you look. *Nat. Med.* 2013;19(5):540-541.
- Nepusz, T., Yu, H.Y. and Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat. Meth.* 2012;9(5):471-U481.
- Neyshabur, B., Khadem, A., Hashemifar, S. and Arab, S.S. NETAL: a new graph-based method for global alignment of protein-protein interaction networks. *Bioinformatics* 2013;29(13):1654-1662.
- Nickel, W. The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. *Eur. J. Biochem.* 2003;270(10):2109-2119.
- Nielsen, H., Brunak, S. and von Heijne, G. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 1999;12(1):3-9.

Nishimura, T., Futami, E., Taneichi, A., Mori, T. and Hattori, A. Decorin expression during development of bovine skeletal muscle and its role in morphogenesis of the intramuscular connective tissue. *Cells Tissues Organs* 2002;171(2-3):199-214.

Novakofski, J. Adipogenesis: usefulness of in vitro and in vivo experimental models. *Journal of animal science* 2004;82(3):905-915.

Nowak, K.J., Wattanasirichaigoon, D., Goebel, H.H., Wilce, M., Pelin, K., Donner, K., Jacob, R.L., Hubner, C., Oexle, K., Anderson, J.R., Verity, C.M., North, K.N., Iannaccone, S.T., Muller, C.R., Nurnberg, P., Muntoni, F., Sewry, C., Hughes, I., Sutphen, R., Lacson, A.G., Swoboda, K.J., Vigneron, J., Wallgren-Pettersson, C., Beggs, A.H. and Laing, N.G. Mutations in the skeletal muscle alpha-actin gene in patients with actin myopathy and nemaline myopathy. *Nature genetics* 1999;23(2):208-212.

O

Ohsaki, H., Sawa, T., Sasazaki, S., Kano, K., Taniguchi, M., Mukai, F. and Mannen, H. Stearoyl-CoA desaturase mRNA expression during bovine adipocyte differentiation in primary culture derived from Japanese Black and Holstein cattle. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 2007;148(3):629-634.

Ohyama, M., Matsuda, K., Torii, S., Matsui, T., Yano, H., Kawada, T. and Ishihara, T. The interaction between vitamin A and thiazolidinedione on bovine adipocyte differentiation in primary culture. *J. Anim. Sci.* 1998;76(1):61-65.

Ondov, B.D., Bergman, N.H. and Phillippy, A.M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 2011;12:385.

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G. and Hermjakob, H. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 2014;42(Database issue):D358-363.

Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkman, F.S., Cesareni, G., Chatr-aryamontri, A., Chautard, E., Chen, C., Dumousseau, M., Goll, J., Hancock, R.E., Hannick, L.I., Jurisica, I., Khadake, J., Lynn, D.J., Mahadevan, U., Perfetto, L., Raghunath, A., Ricard-Blum, S., Roechert, B., Salwinski, L., Stumpflen, V., Tyers, M., Uetz, P., Xenarios, I. and Hermjakob, H. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nature methods* 2012;9(4):345-350.

Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stumpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., Salama, J.J., Moore, S., Wojcik, J., Bader, G.D., Vidal, M., Cusick, M.E., Gerstein, M., Gavin, A.C., Superti-Furga, G., Greenblatt, J., Bader, J., Uetz, P., Tyers, M., Legrain, P., Fields, S., Mulder, N., Gilson, M., Niepmann, M., Burgoon, L., De Las Rivas, J., Prieto, C., Perreau, V.M., Hogue, C., Mewes, H.W., Apweiler, R., Xenarios, I., Eisenberg, D., Cesareni, G. and Hermjakob, H. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nature biotechnology* 2007;25(8):894-898.

Otto, A., Schmidt, C., Luke, G., Allen, S., Valasek, P., Muntoni, F., Lawrence-Watt, D. and Patel, K. Canonical Wnt signalling induces satellite-cell proliferation during adult skeletal muscle regeneration. *J. Cell Sci.* 2008;121(17):2939-2950.

P

Pallero, M.A., Elzie, C.A., Chen, J.P., Mosher, D.F. and Murphy-Ullrich, J.E. Thrombospondin 1 binding to calreticulin-LRP1 signals resistance to anoikis. *FASEB J.* 2008;22(11):3968-3979.

- Park, A., Kim, W.K. and Bae, K.H. Distinction of white, beige and brown adipocytes derived from mesenchymal stem cells. *World journal of stem cells* 2014;6(1):33-42.
- Park, K.H., Kang, J.W., Lee, E.M., Kim, J.S., Rhee, Y.H., Kim, M., Jeong, S.J., Park, Y.G. and Kim, S.H. Melatonin promotes osteoblastic differentiation through the BMP/ERK/Wnt signaling pathways. *J. Pineal Res.* 2011;51(2):187-194.
- Patapoutian, A., Yoon, J.K., Miner, J.H., Wang, S., Stark, K. and Wold, B. Disruption of the mouse MRF4 gene identifies multiple waves of myogenesis in the myotome. *Development* 1995;121(10):3347-3358.
- Pavlou, M.P. and Diamandis, E.P. The cancer cell secretome: a good source for discovering biomarkers? *Journal of proteomics* 2010;73(10):1896-1906.
- Pearson, W.R. An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* 2013;Chapter 3:Unit3 1.
- Pedersen, B.K. Muscles and their myokines. *The Journal of experimental biology* 2011;214(Pt 2):337-346.
- Pedersen, B.K., Akerstrom, T.C., Nielsen, A.R. and Fischer, C.P. Role of myokines in exercise and metabolism. *J Appl Physiol (1985)* 2007;103(3):1093-1098.
- Pedersen, B.K. and Fischer, C.P. Beneficial health effects of exercise--the role of IL-6 as a myokine. *Trends Pharmacol. Sci.* 2007;28(4):152-156.
- Perry, R.L. and Rudnick, M.A. Molecular mechanisms regulating myogenic determination and differentiation. *Front. Biosci.* 2000;5:D750-767.
- Persico, M., Ceol, A., Gavrila, C., Hoffmann, R., Florio, A. and Cesareni, G. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* 2005;6 Suppl 4:S21.
- Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 2011;8(10):785-786.
- Pethick, D.W. and Dunshea, F.R. The Partitioning of Fat in Farm Animals. *Proceedings of the Nutrition Society of Australia* 1996;20: 3-13.
- Petrovic, N., Walden, T.B., Shabalina, I.G., Timmons, J.A., Cannon, B. and Nedergaard, J. Chronic peroxisome proliferator-activated receptor gamma (PPARgamma) activation of epididymally derived white adipocyte cultures reveals a population of thermogenically competent, UCP1-containing adipocytes molecularly distinct from classic brown adipocytes. *The Journal of biological chemistry* 2010;285(10):7153-7164.
- Picard, B., Barboiron, C., Kaspric, N., Vonnahme, K. and Bonnet, M. Impact on muscle fiber properties of maternal nutrient restriction followed by realimentation from early to mid-gestation in beef cows. *61st International Congress of Meat Science and Technology* 2015.
- Picard, B., Berri, C., Lefaucheur, L., Molette, C., Sayd, T. and Terlouw, C. Skeletal muscle proteomics in livestock production. *Briefings in functional genomics* 2010;9(3):259-278.
- Picard, B. and Cassar-Malek, I. Evidence for expression of IIb myosin heavy chain isoform in some skeletal muscles of Blonde d'Aquitaine bulls. *Meat science* 2009;82(1):30-36.
- Picard, B., Cassar-Malek, I., Guillemin, N. and Bonnet, M. Quest for Novel Muscle Pathway Biomarkers by Proteomics in Beef Production. In: Moo-Young, M., editor, *Comprehensive Biotechnology (Second Edition)*. Burlington: Academic Press; 2011. p. 395-405.
- Picard, B., Duris, M.P. and Jurie, C. Classification of bovine muscle fibres by different histochemical techniques. *Histochem. J.* 1998;30(7):473-479.
- Picard, B., Jurie, C., Cassar-Malek, I. and Hocquette, J.F. Typologie et ontogenèse des fibres musculaires chez le bovin. *INRA Productions Animales* 2003;16(2):125-131.
- Picard, B., Lefaucheur, L., Berri, C. and Duclos, M.J. Muscle fibre ontogenesis in farm animal species. *Reprod. Nutr. Dev.* 2002;42(5):415-431.
- Picard, B., Robelin, J., Pons, F. and Geay, Y. Comparison of the foetal development of fibre types in four bovine muscles. *J. Muscle Res. Cell Motil.* 1994;15(4):473-486.

Pisani, D.F., Djedaini, M., Beranger, G.E., Elabd, C., Scheideler, M., Ailhaud, G. and Amri, E.Z. Differentiation of Human Adipose-Derived Stem Cells into "Brite" (Brown-in-White) Adipocytes. *Frontiers in endocrinology* 2011;2:87.

Poirier, F., Timmons, P.M., Chan, C.T., Guenet, J.L. and Rigby, P.W. Expression of the L14 lectin during mouse embryogenesis suggests multiple roles during pre- and post-implantation development. *Development* 1992;115(1):143-155.

Pope, M., Budge, H. and Symonds, M.E. The developmental transition of ovine adipose tissue through early life. *Acta Physiol (Oxf)* 2014;210(1):20-30.

Poulos, S.P., Hausman, D.B. and Hausman, G.J. The development and endocrine functions of adipose tissue. *Mol. Cell. Endocrinol.* 2010;323(1):20-34.

Prassas, I., Chrystoja, C.C., Makawita, S. and Diamandis, E.P. Bioinformatic identification of proteins with tissue-specific expression for biomarker discovery. *BMC medicine* 2012;10:39.

Prieto, C. and De Las Rivas, J. APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res* 2006;34(Web Server issue):W298-302.

Purroy, A., Mendizabal, J. Soret, B., Arana, A. et Mendizabal, F;. Changes in cell number and size and in lipogenic enzyme activity in adipose tissues during growth and fattening of lacha (manech) lamb. *Annales de zootechnie* 1997;46:309-319.

Q-R

Quadro, L., Hamberger, L., Gottesman, M.E., Wang, F.W., Colantuoni, V., Blaner, W.S. and Mendelsohn, C.L. Pathways of vitamin A delivery to the embryo: Insights from a new tunable model of embryonic vitamin A deficiency. *Endocrinology* 2005;146(10):4479-4490.

Ramirez, F., Lawyer, G. and Albrecht, M. Novel search method for the discovery of functional relationships. *Bioinformatics* 2012;28(2):269-276.

Randle, P.J. Anaerobic uptake of glucose in vitro by the isolated rat diaphragm. *Nature* 1956;178(4540):983-984.

Rao, V.S., Srinivas, K., Sujini, G.N. and Kumar, G.N. Protein-protein interaction detection: methods and analysis. *International journal of proteomics* 2014;2014:147648.

Raschke, S. and Eckel, J. Adipo-myokines: two sides of the same coin--mediators of inflammation and mediators of exercise. *Mediators Inflamm.* 2013;2013:320724.

Raschke, S., Elsen, M., Gassenhuber, H., Sommerfeld, M., Schwahn, U., Brockmann, B., Jung, R., Wisloff, U., Tjonna, A.E., Raastad, T., Hallen, J., Norheim, F., Drevon, C.A., Romacho, T., Eckardt, K. and Eckel, J. Evidence against a beneficial effect of irisin in humans. *PloS one* 2013;8(9):e73680.

Rawls, A., Valdez, M.R., Zhang, W., Richardson, J., Klein, W.H. and Olson, E.N. Overlapping functions of the myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice. *Development* 1998;125(13):2349-2358.

Reese, J.T., Childers, C.P., Sundaram, J.P., Dickens, C.M., Childs, K.L., Vile, D.C. and Elsik, C.G. Bovine Genome Database: supporting community annotation and analysis of the *Bos taurus* genome. *BMC genomics* 2010;11:645.

Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* 2007;35(Web Server issue):W193-200.

Renaud, Y., Baillif, A., Perez, J.B., Agier, M., Mephu Nguifo, E. and Mirouse, V. DroPNet: a web portal for integrated analysis of *Drosophila* protein-protein interaction networks. *Nucleic Acids Res* 2012;40(Web Server issue):W134-139.

Rice, P., Longden, I. and Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG* 2000;16(6):276-277.

Rivenbark, A.G., O'Connor, S.M. and Coleman, W.B. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *Am. J. Pathol.* 2013;183(4):1113-1124.

- Robelin, J. Growth of Adipose Tissues in Cattle - Partitioning between Depots, Chemical-Composition and Cellularity - a Review. *Livest. Prod. Sci.* 1986;14(4):349-364.
- Robelin, J., Picard, B., Listrat, A., Jurie, C., Barboiron, C., Pons, F. and Geay, Y. Myosin expression in semitendinosus muscle during fetal development of cattle: immunocytochemical and electrophoretic analyses. *Reprod. Nutr. Dev.* 1993;33(1):25-41.
- Robelin, J.e.C., L. Differentiation, growth and development of adipose tissue. *Productions animales* 1990;3:243-252.
- Robinson, P., Lipscomb, S., Preston, L.C., Altin, E., Watkins, H., Ashley, C.C. and Redwood, C.S. Mutations in fast skeletal troponin I, troponin T, and beta-tropomyosin that cause distal arthrogyrosis all increase contractile function. *FASEB J.* 2007;21(3):896-905.
- Rodriguez, J., Vernus, B., Chelh, I., Cassar-Malek, I., Gabillard, J.C., Sassi, A.H., Seiliez, I., Picard, B. and Bonnieu, A. Myostatin and the skeletal muscle atrophy and hypertrophy signaling pathways. *Cell. Mol. Life Sci.* 2014;71(22):4361-4371.
- Romacho, T., Elsen, M., Rohrborn, D. and Eckel, J. Adipose tissue and its role in organ crosstalk. *Acta Physiol (Oxf)* 2014;210(4):733-753.
- Romao, J.M., He, M.L., McAllister, T.A. and Guan, L.L. Effect of age on bovine subcutaneous fat proteome: Molecular mechanisms of physiological variations during beef cattle growth. *J. Anim. Sci.* 2014;92(8):3316-3327.
- Ronnov-Jessen, L., Villadsen, R., Edwards, J.C. and Petersen, O.W. Differential expression of a chloride intracellular channel gene, CLIC4, in transforming growth factor-beta1-mediated conversion of fibroblasts to myofibroblasts. *Am. J. Pathol.* 2002;161(2):471-480.
- Rosen, E.D. and MacDougald, O.A. Adipocyte differentiation from the inside out. *Nat. Rev. Mol. Cell Biol.* 2006;7(12):885-896.
- Rubin, B.Y., Anderson, S.L., Xing, L., Powell, R.J. and Tate, W.P. Interferon induces tryptophanyl-tRNA synthetase expression in human fibroblasts. *The Journal of biological chemistry* 1991;266(36):24245-24248.
- Rudnicki, M.A., Braun, T., Hinuma, S. and Jaenisch, R. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* 1992;71(3):383-390.
- Rudnicki, M.A., Schnegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H. and Jaenisch, R. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* 1993;75(7):1351-1359.

S

- Sable, R. and Jois, S. Surfing the Protein-Protein Interaction Surface Using Docking Methods: Application to the Design of PPI Inhibitors. *Molecules* 2015;20(6):11569-11603.
- Sadri, H., Mielenz, M., Morel, I., Bruckmaier, R.M. and van Dorland, H.A. Plasma leptin and mRNA expression of lipogenesis and lipolysis-related factors in bovine adipose tissue around parturition. *J. Anim. Physiol. Anim. Nutr.* 2011;95(6):790-797.
- Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004;32(Database issue):D449-451.
- Sanchez-Gurmaches, J. and Guertin, D.A. Adipocyte lineages: tracing back the origins of fat. *Biochim. Biophys. Acta* 2014;1842(3):340-351.
- Sanchez-Gurmaches, J. and Guertin, D.A. Adipocytes arise from multiple lineages that are heterogeneously and dynamically distributed. *Nature communications* 2014;5:4099.
- Schatz, M.C. and Langmead, B. The DNA Data Deluge: Fast, efficient genome sequencing machines are spewing out more data than geneticists can analyze. *IEEE spectrum* 2013;50(7):26-33.
- Scheideler, M., Elabd, C., Zaragosi, L.E., Chiellini, C., Hackl, H., Sanchez-Cabo, F., Yadav, S., Duszka, K., Friedl, G., Papak, C., Prokesch, A., Windhager, R., Ailhaud, G., Dani, C., Amri, E.Z. and Trajanoski, Z. Comparative transcriptomics of human multipotent stem cells during adipogenesis and osteoblastogenesis. *BMC genomics* 2008;9:340.

- Schiaffino, S. and Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol. Rev.* 2011;91(4):1447-1531.
- Schmidt, A., Forne, I. and Imhof, A. Bioinformatic analysis of proteomics data. *BMC Syst. Biol.* 2014;8 Suppl 2:S3.
- Seale, P., Bjork, B., Yang, W.L., Kajimura, S., Chin, S., Kuang, S.H., Scime, A., Devarakonda, S., Conroe, H.M., Erdjument-Bromage, H., Tempst, P., Rudnicki, M.A., Beier, D.R. and Spiegelman, B.M. PRDM16 controls a brown fat/skeletal muscle switch. *Nature* 2008;454(7207):961-U927.
- Seale, P., Kajimura, S. and Spiegelman, B.M. Transcriptional control of brown adipocyte development and physiological function--of mice and men. *Genes Dev.* 2009;23(7):788-797.
- Seldin, M.M., Peterson, J.M., Byerly, M.S., Wei, Z.K. and Wong, G.W. Myonectin (CTRP15), a Novel Myokine That Links Skeletal Muscle to Systemic Lipid Homeostasis. *J. Biol. Chem.* 2012;287(15):11968-11980.
- Seo, Y., Lee, K., Park, K., Bae, K. and Choi, I. A proteomic assessment of muscle contractile alterations during unloading and reloading. *J Biochem* 2006;139(1):71-80.
- Sillence, M.N. Technologies for the Control of Fat and Lean Deposition in Livestock. *Vet. J.* 2004;167(3):242-257.
- Skalnikova, H., Motlik, J., Gadher, S.J. and Kovarova, H. Mapping of the secretome of primary isolates of mammalian cells, stem cells and derived cell lines. *Proteomics* 2011;11(4):691-708.
- Skop, V., Cahova, M., Dankova, H., Papackova, Z., Palenickova, E., Svoboda, P., Zidkova, J. and Kazdova, L. Autophagy inhibition in early but not in later stages prevents 3T3-L1 differentiation: Effect on mitochondrial remodeling. *Differentiation* 2014.
- Skunca, N., Altenhoff, A. and Dessimoz, C. Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.* 2012;8(5):e1002533.
- Sleator, R.D. and Walsh, P. An overview of in silico protein function prediction. *Arch. Microbiol.* 2010;192(3):151-155.
- Smith, S.B., Carstens, G.E., Randel, R.D., Mersmann, H.J. and Lunt, D.K. Brown adipose tissue development and metabolism in ruminants. *Journal of animal science* 2004;82(3):942-954.
- Smith, S.B., Lunt, D.K., Chung, K.Y., Choi, C.B., Tume, R.K. and Zembayashi, M. Adiposity, fatty acid composition, and delta-9 desaturase activity during growth in beef cattle. *Anim. Sci. J.* 2006;77(5):478-486.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. and Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;27(3):431-432.
- Soret, B., Lee, H.J., Finley, E., Lee, S.C. and Vernon, R.G. Regulation of differentiation of sheep subcutaneous and abdominal preadipocytes in culture. *J. Endocrinol.* 1999;161(3):517-524.
- Spinelli, L., Gambette, P., Chapple, C.E., Robisson, B., Baudot, A., Garreta, H., Tichit, L., Guenoche, A. and Brun, C. Clust&See: a Cytoscape plugin for the identification, visualization and manipulation of network clusters. *Biosystems* 2013;113(2):91-95.
- Sprenger, J., Fink, J.L., Karunaratne, S., Hanson, K., Hamilton, N.A. and Teasdale, R.D. LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res.* 2008;36:D230-D233.
- Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H., Lehvaslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D. and Birney, E. The Bioperl toolkit: Perl modules for the life sciences. *Genome research* 2002;12(10):1611-1618.
- Steensberg, A., van Hall, G., Osada, T., Sacchetti, M., Saltin, B. and Klarlund Pedersen, B. Production of interleukin-6 in contracting human skeletal muscles can account for the exercise-induced increase in plasma interleukin-6. *J. Physiol.* 2000;529 Pt 1:237-242.
- Stockinger, H., Altenhoff, A.M., Arnold, K., Bairoch, A., Bastian, F., Bergmann, S., Bougueleret, L., Bucher, P., Delorenzi, M., Lane, L., Le Mercier, P., Lisacek, F., Michielin, O., Palagi, P.M., Rougemont, J., Schwede, T., von Mering, C., van Nimwegen, E., Walther, D., Xenarios, I., Zavolan, M., Zdobnov, E.M., Zoete, V. and Appel, R.D. Fifteen years SIB Swiss Institute of Bioinformatics: life science databases, tools and support. *Nucleic Acids Res* 2014;42(Web Server issue):W436-441.

Sudre, K., Leroux, C., Pietu, G., Cassar-Malek, I., Petit, E., Listrat, A., Auffray, C., Picard, B., Martin, P. and Hocquette, J.F. Transcriptome analysis of two bovine muscles during ontogenesis. *J. Biochem. (Tokyo)* 2003;133(6):745-756.

Sun, D.B., Zhang, H., Guo, D.H., Sun, A.G. and Wang, H.B. Shotgun Proteomic Analysis of Plasma from Dairy Cattle Suffering from Footrot: Characterization of Potential Disease-Associated Factors. *PLoS one* 2013;8(2).

Sun, Z., Cheng, Z., Taylor, C.A., McConkey, B.J. and Thompson, J.E. Apoptosis induction by eIF5A1 involves activation of the intrinsic mitochondrial pathway. *Journal of cellular physiology* 2010;223(3):798-809.

Suzuki, K., Kishioka, Y., Wakamatsu, J. and Nishimura, T. Decorin activates Akt downstream of IGF-IR and promotes myoblast differentiation. *Anim. Sci. J.* 2013;84(9):669-674.

I

Taga, H.: Université européenne de Bretagne; 2011. Ontogénèse des tissus adipeux chez le foetus bovin: identification des processus cellulaires et moléculaires sous-jacents à l'adipogénèse.

Taga, H., Bonnet, M., Picard, B., Zingaretti, M.C., Cassar-Malek, I., Cinti, S. and Chilliard, Y. Adipocyte metabolism and cellularity are related to differences in adipose tissue maturity between Holstein and Charolais or Blond d'Aquitaine fetuses. *Journal of animal science* 2011;89(3):711-721.

Taga, H., Chilliard, Y., Meunier, B., Chambon, C., Picard, B., Zingaretti, M.C., Cinti, S. and Bonnet, M. Cellular and molecular large-scale features of fetal adipose tissue: is bovine perirenal adipose tissue brown? *Journal of cellular physiology* 2012;227(4):1688-1700.

Taga, H., Chilliard, Y., Picard, B., Zingaretti, M.C. and Bonnet, M. Foetal bovine intermuscular adipose tissue exhibits histological and metabolic features of brown and white adipocytes during the last third of pregnancy. *Animal : an international journal of animal bioscience* 2012;6(4):641-649.

Tan, S.H., Reverter, A., Wang, Y., Byrne, K.A., McWilliam, S.M. and Lehnert, S.A. Gene expression profiling of bovine in vitro adipogenesis using a cDNA microarray. *Funct. Integr. Genomics* 2006;6(3):235-249.

Tang, Q.Q. and Lane, M.D. Adipogenesis: from stem cell to adipocyte. *Annu. Rev. Biochem.* 2012;81:715-736.

Taniguchi, M., Guan, L.L., Zhang, B., Dodson, M.V., Okine, E. and Moore, S.S. Gene expression patterns of bovine perimuscular preadipocytes during adipogenesis. *Biochem. Biophys. Res. Commun.* 2008;366(2):346-351.

Tarcea, V.G., Weymouth, T., Ade, A., Bookvich, A., Gao, J., Mahavisno, V., Wright, Z., Chapman, A., Jayapandian, M., Ozgur, A., Tian, Y., Cavalcoli, J., Mirel, B., Patel, J., Radev, D., Athey, B., States, D. and Jagadish, H.V. Michigan molecular interactions r2: from interacting proteins to pathways. *Nucleic Acids Res* 2009;37(Database issue):D642-646.

Timmons, J.A., Wennmalm, K., Larsson, O., Walden, T.B., Lassmann, T., Petrovic, N., Hamilton, D.L., Gimeno, R.E., Wahlestedt, C., Baar, K., Nedergaard, J. and Cannon, B. Myogenic gene expression signature establishes that brown and white adipocytes originate from distinct cell lineages. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(11):4401-4406.

Tore, O. Structure and Evolution of Weighted Networks. 2009.

Torii, S.-I., Ohyama, M., Matsui, T. and Yano, H. Ascorbic acid-2-phosphate enhances adipocyte differentiation of cultured stromal vascular cells prepared from bovine perirenal adipose tissue. *Anim. Sci. Technol.* 1998;69(5):439-444.

Tortora, G.J. and Grabowski, S.R. Principes d'anatomie et de physiologie. 1994.

Tournayre, J., I., C.-M., Reichstadt, M., B., P., Kaspric, N. and Bonnet, M. Fat&MuscleDB: integrating 'omics' data from adipose tissue and muscle. In, *Farm Animal Proteomics COST Action 1002*. Milan; 2014. p. 91-94.

Tournayre, J., I., C.-M., Reichstadt, M., B., P., Kaspric, N. and Bonnet, M. Fat&MuscleDB: a database to understand tissue growth processes contributing to body or muscle composition. In, *JOBIM: Journée Ouvertes en Biologie, Informatique et Mathématiques*. Clermont-Ferrand; 2015. p. 14.

Trayhurn, P., Bing, C. and Wood, I.S. Adipose tissue and adipokines--energy regulation from the human perspective. *J. Nutr.* 2006;136(7 Suppl):1935S-1939S.

Trayhurn, P., Drevon, C.A. and Eckel, J. Secreted proteins from adipose tissue and skeletal muscle - adipokines, myokines and adipose/muscle cross-talk. *Arch. Physiol. Biochem.* 2011;117(2):47-56.

Trayhurn, P., Temple, N.J. and Van Aerde, J. Evidence from immunoblotting studies on uncoupling protein that brown adipose tissue is not present in the domestic pig. *Can. J. Physiol. Pharmacol.* 1989;67(12):1480-1485.

Tseng, Y.H., Kokkotou, E., Schulz, T.J., Huang, T.L., Winnay, J.N., Taniguchi, C.M., Tran, T.T., Suzuki, R., Espinoza, D.O., Yamamoto, Y., Ahrens, M.J., Dudley, A.T., Norris, A.W., Kulkarni, R.N. and Kahn, C.R. New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature* 2008;454(7207):1000-1004.

Tseng, Y.H., Kokkotou, E., Schulz, T.J., Huang, T.L., Winnay, J.N., Taniguchi, C.M., Tran, T.T., Suzuki, R., Espinoza, D.O., Yamamoto, Y., Ahrens, M.J., Dudley, A.T., Norris, A.W., Kulkarni, R.N. and Kahn, C.R. New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature* 2008;454(7207):1000-U1044.

Tzen, C.Y. and Huang, Y.W. Cloning of murine early quiescence-1 gene: the murine counterpart of dermatopontin gene can induce and be induced by cell quiescence. *Exp. Cell Res.* 2004;294(1):30-38.

U-V

Ulitsky, I., Maron-Katz, A., Shavit, S., Sagir, D., Linhart, C., Elkon, R., Tanay, A., Sharan, R., Shiloh, Y. and Shamir, R. Expander: from expression microarrays to networks and functions. *Nat Protoc* 2010;5(2):303-322.

UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* 2015;43(Database issue):D204-212.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 2012;40(15):e115.

Vasiliou, V., Pappa, A. and Petersen, D.R. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chem. Biol. Interact.* 2000;129(1-2):1-19.

Vernon, R.G. The growth and metabolism of adipocytes. In: Butterworths, editor, *Control and manipulation of animal growth*. Buttery P.J., Lindsay D.B., Haynes N.B.; 1986. p. 67-83.

Verovskaya, E., Broekhuis, M.J., Zwart, E., Ritsema, M., van Os, R., de Haan, G. and Bystrykh, L.V. Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. *Blood* 2013;122(4):523-532.

Vickers, K.C., Roteta, L.A., Hucheson-Dilks, H., Han, L. and Guo, Y. Mining diverse small RNA species in the deep transcriptome. *Trends Biochem. Sci.* 2015;40(1):4-7.

Vivian, J.L., Gan, L., Olson, E.N. and Klein, W.H. A hypomorphic myogenin allele reveals distinct myogenin expression levels required for viability, skeletal muscle development, and sternum formation. *Dev. Biol.* 1999;208(1):44-55.

W

Wang, S., Zhang, B. and Faller, D.V. Prohibitin requires Brg-1 and Brm for the repression of E2F and cell growth. *EMBO J.* 2002;21(12):3019-3028.

- Wanichthanarak, K., Fahrman, J.F. and Grapov, D. Genomic, Proteomic, and Metabolomic Data Integration Strategies. *Biomarker insights* 2015;10(Suppl 4):1-6.
- Weigert, C., Lehmann, R., Hartwig, S. and Lehr, S. The secretome of the working human skeletal muscle--a promising opportunity to combat the metabolic disaster? *Proteomics. Clinical applications* 2014;8(1-2):5-18.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S. and et al. The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* 1991;251(4995):761-766.
- Wendler, C.C., Schmoltdt, A., Flentke, G.R., Case, L.C., Quadro, L., Blaner, W.S., Lough, J. and Smith, S.M. Increased fibronectin deposition in embryonic hearts of retinol-binding protein-null mice. *Circ. Res.* 2003;92(8):920-928.
- Wigmore, P.M. and Evans, D.J. Molecular and cellular mechanisms involved in the generation of fiber diversity during myogenesis. *Int. Rev. Cytol.* 2002;216:175-232.
- Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.H., Bantscheff, M., Gerstmair, A., Faerber, F. and Kuster, B. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509(7502):582-587.
- Wilson, P.A., Gardner, S.D., Lambie, N.M., Commans, S.A. and Crowther, D.J. Characterization of the human patatin-like phospholipase family. *J. Lipid Res.* 2006;47(9):1940-1949.
- Wilson, S.J., McEwan, J.C., Sheard, P.W. and Harris, A.J. Early stages of myogenesis in a large mammal: formation of successive generations of myotubes in sheep tibialis cranialis muscle. *J. Muscle Res. Cell Motil.* 1992;13(5):534-550.
- Woelders, H., Te Pas, M.F., Bannink, A., Veerkamp, R.F. and Smits, M.A. Systems biology in animal sciences. *Animal : an international journal of animal bioscience* 2011;5(7):1036-1047.
- Wolfing, B., Buechler, C., Weigert, J., Neumeier, M., Aslanidis, C., Schoelmerich, J. and Schaffler, A. Effects of the new C1q/TNF-related protein (CTRP-3) "cartonectin" on the adipocytic secretion of adipokines. *Obesity (Silver Spring)* 2008;16(7):1481-1486.
- Woronuk, G.N., Marquess, F.L., James, S.T., Palmer, J., Berryere, T., Deobald, H., Howie, S. and Kononoff, P.J. Association of leptin genotypes with beef cattle characteristics. *Anim. Genet.* 2012;43(5):608-610.
- Wu, J., Bostrom, P., Sparks, L.M., Ye, L., Choi, J.H., Giang, A.H., Khandekar, M., Virtanen, K.A., Nuutila, P., Schaart, G., Huang, K., Tu, H., van Marken Lichtenbelt, W.D., Hoeks, J., Enerback, S., Schrauwen, P. and Spiegelman, B.M. Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. *Cell* 2012;150(2):366-376.

X-Y-Z

- Yablonka-Reuveni, Z., Rudnicki, M.A., Rivera, A.J., Primig, M., Anderson, J.E. and Natanson, P. The transition from proliferation to differentiation is delayed in satellite cells from mice lacking MyoD. *Dev. Biol.* 1999;210(2):440-455.
- Yamauchi, T., Kamon, J., Minokoshi, Y., Ito, Y., Waki, H., Uchida, S., Yamashita, S., Noda, M., Kita, S., Ueki, K., Eto, K., Akanuma, Y., Froguel, P., Foufelle, F., Ferre, P., Carling, D., Kimura, S., Nagai, R., Kahn, B.B. and Kadowaki, T. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nat. Med.* 2002;8(11):1288-1295.
- Yan, X., Zhu, M.J., Dodson, M.V. and Du, M. Developmental programming of fetal skeletal muscle and adipose tissue development. *Journal of genomics* 2013;1:29-38.
- Yokota, S., Yanagi, H., Yura, T. and Kubota, H. Cytosolic chaperonin is up-regulated during cell growth - Preferential, expression and binding to tubulin at G(1)/S transition through early S phase. *J. Biol. Chem.* 1999;274(52):37070-37078.

- Yokota, T., Meka, C.S., Medina, K.L., Igarashi, H., Comp, P.C., Takahashi, M., Nishida, M., Oritani, K., Miyagawa, J., Funahashi, T., Tomiyama, Y., Matsuzawa, Y. and Kincade, P.W. Paracrine regulation of fat cell formation in bone marrow cultures via adiponectin and prostaglandins. *J. Clin. Invest.* 2002;109(10):1303-1310.
- Yonekura, S., Tokutake, Y., Hirota, S., Rose, M.T., Katoh, K. and Aso, H. Proliferating bovine intramuscular preadipocyte cells synthesize leptin. *Domest. Anim. Endocrinol.* 2013;45(1):33-37.
- Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 2010;26(13):1608-1615.
- Yu, Y.H., Chang, Y.C., Su, T.H., Nong, J.Y., Li, C.C. and Chuang, L.M. Prostaglandin reductase-3 negatively modulates adipogenesis through regulation of PPAR gamma activity. *J. Lipid Res.* 2013;54(9):2391-2399.
- Yuryev, A., Kotelnikova, E. and Daraselia, N. Ariadne's ChemEffect and Pathway Studio knowledge base. *Expert opinion on drug discovery* 2009;4(12):1307-1318.
- Zaragosi, L.E., Wdziekonski, B., Villageois, P., Keophiphath, M., Maumus, M., Tchkonja, T., Bourlier, V., Mohsen-Kanson, T., Ladoux, A., Elabd, C., Scheideler, M., Trajanoski, Z., Takashima, Y., Amri, E.Z., Lacasa, D., Sengenès, C., Ailhaud, G., Clement, K., Bouloumie, A., Kirkland, J.L. and Dani, C. Activin A Plays a Critical Role in Proliferation and Differentiation of Human Adipose Progenitors. *Diabetes* 2010;59(10):2513-2521.
- Zaragosi, L.E., Wdziekonski, B., Villageois, P., Keophiphath, M., Maumus, M., Tchkonja, T., Bourlier, V., Mohsen-Kanson, T., Ladoux, A., Elabd, C., Scheideler, M., Trajanoski, Z., Takashima, Y., Amri, E.Z., Lacasa, D., Sengenès, C., Ailhaud, G., Clement, K., Bouloumie, A., Kirkland, J.L. and Dani, C. Activin a plays a critical role in proliferation and differentiation of human adipose progenitors. *Diabetes* 2010;59(10):2513-2521.
- Zhang, W., Behringer, R.R. and Olson, E.N. Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies. *Genes Dev.* 1995;9(11):1388-1399.
- Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassel, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A. and Salzberg, S.L. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009;10(4).
- Zou, D., Ma, L., Yu, J. and Zhang, Z. Biological databases for human research. *Genomics Proteomics Bioinformatics* 2015;13(1):55-63.