



**HAL**  
open science

## Exploitation bioinformatique de la comparaison des transcriptomes de tissus adipeux de chèvres lactantes et non lactantes

Yannick Gestin, Yannick Faulconnier, Céline Boby, Christine Leroux

### ► To cite this version:

Yannick Gestin, Yannick Faulconnier, Céline Boby, Christine Leroux. Exploitation bioinformatique de la comparaison des transcriptomes de tissus adipeux de chèvres lactantes et non lactantes. Sciences du Vivant [q-bio]. 2016. hal-04059020

**HAL Id: hal-04059020**

**<https://hal.inrae.fr/hal-04059020v1>**

Submitted on 5 Apr 2023

**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.



Distributed under a Creative Commons Attribution 4.0 International License

Université d'Auvergne - IUT de Clermont Ferrand –  
Campus universitaire d'Aurillac - Département Génie Biologique

**Yannick GESTIN**

2<sup>ème</sup> Année 2015/2016

DUT Génie Biologique -BioInformatique

---

# Exploitation bioinformatique de la comparaison des transcriptomes de tissus adipeux de chèvres lactantes et non lactantes

---

Dans le cadre d'un stage effectuée du 4 avril au 10 juin 2016



**Encadrantes :**

Yannick FAULCONNIER

Céline BOBY

Institut Nationale de la **R**echerche **A**gronomique (**INRA**)  
Unité Mixte de **R**echerche sur les **H**erbivores (**UMRH**)  
Centre de Recherches Auvergne-Rhône-Alpes  
63122 Saint-Genès-Champanelle

## REMERCIEMENTS

Je commencerais par remercier Madame Isabelle Veissier, directrice de l'UMRH, de m'avoir accueilli pour la durée de mon stage au sein de son unité à l'Inra de Theix.

Mes plus profonds remerciements vont à Céline Boby ingénieure d'étude et Yannick Faulconnier ingénieure de recherche, mes supers co-encadrantes qui m'ont guidé durant ces deux mois. Je les remercie de m'avoir choisi pour réaliser ce stage et pour m'avoir aussi bien accueilli dans l'équipe. Mais aussi pour leurs conseils, leur motivation et pour le temps qu'elles m'ont accordé tout au long de la rédaction de rapport.

Je tiens également à remercier Muriel Bonnet, responsable d'équipe Biomarqueurs, pour m'avoir accueillie dans l'équipe ainsi que pour ses conseils judicieux dans l'utilisation d'outils bioinformatiques.

Christine Leroux, directrice de recherche, pour sa relecture minutieuse de mon rapport ainsi que pour ses conseils tout au long de mon stage.

Anne de la Foye, ingénieure d'étude, pour sa patience et sa disponibilité lorsque j'avais besoin d'une aide statistique.

Jérémy Tournayre et Jordann Domagalski pour être venus à mon secours à plusieurs reprises lors de mes bidouillages informatiques.

Karol Pawloski, pour avoir partagé son bureau et ses conseils, tout en supportant mes exclamations de joies et de colères.

Et le mot de la fin revient aux personnes des équipes Biomarqueurs et Peraq pour leur accueil chaleureux et pour tous les moments partagés, que ce soit, les sorties au restaurant, les pauses cafés ou au self. Merci pour ces conversations, ces éclats de rires et ces petits taquinages. Le minot vous souhaite de ne jamais perdre cette bonne humeur et cet esprit bon enfant.

<b>REMERCIEMENTS .....</b>	
<b>RESUME .....</b>	
<b>ABSTRACT.....</b>	
<b>INTRODUCTION .....</b>	<b>1</b>
<b>1. STRUCTURE D'ACCUEIL .....</b>	<b>2</b>
1.1. L'Institut National de la Recherche Agronomique (Inra).....	2
1.2. Le centre de recherche Auvergne-Rhône-Alpes .....	3
1.3. L'Unité Mixte de Recherche sur les Herbivores (UMRH) .....	4
1.4. L'équipe Biomarqueurs (BIOM).....	5
<b>2. ETUDE BIBLIOGRAPHIQUE DU SUJET.....</b>	<b>6</b>
2.1. Généralités sur le Tissu Adipeux (TA) .....	6
2.1.1. Lipogenèse.....	7
2.1.2. Lipolyse .....	8
2.1.3. Cycles lipogenèse/lipolyse.....	9
2.2. Métabolisme du tissu adipeux chez le ruminant en lactation.....	9
<b>3. MATERIEL ET METHODE.....</b>	<b>11</b>
3.1. Le projet MilkChEST .....	11
3.2. Etude expérimentale .....	12
3.3. Le séquençage Illumina .....	13
3.4. Analyse bio-informatique des données .....	16
3.4.1. ProteINSIDE.....	16
3.4.2. Protein Analysis Through Evolutionary Relationships (PANTHER).....	17

3.4.3.	KEGG (Kyoto Encyclopedia of Genes and Genomes).....	19
3.4.4.	Pathway Studio .....	19
<b>4.</b>	<b>RESULTATS ET DISCUSSION .....</b>	<b>20</b>
<b>4.1.</b>	<b>Identification des gènes.....</b>	<b>20</b>
<b>4.2.</b>	<b>Ontologie des gènes .....</b>	<b>22</b>
4.2.1.	Classification des gènes par GO .....	22
4.2.2.	Enrichissement en GO .....	24
<b>4.3.</b>	<b>Analyse des voies métaboliques (Pathways).....</b>	<b>27</b>
<b>4.4.</b>	<b>Analyse et construction de réseaux.....</b>	<b>27</b>
<b>4.5.</b>	<b>Discussion .....</b>	<b>29</b>
	<b>CONCLUSION .....</b>	<b>32</b>
	<b>BIBLIOGRAPHIE.....</b>	
	<b>TABLE DES SIGLES ET/OU DES ABREVIATIONS.....</b>	
	<b>TABLE DES TABLEAUX ET DES FIGURES.....</b>	
	<b>TABLES DES ANNEXES ET ANNEXES .....</b>	

## RESUME

L'objectif de cette étude a été de comparer les **mécanismes de stockage lipidique** dans les **tissus adipeux de chèvres lactantes** (en phase de mobilisation des réserves lipidiques) et non lactantes (en phase de stockage lipidique). Un séquençage haut débit des ARN (**RNA-seq**) extraits du tissu adipeux a montré que 252 gènes étaient différentiellement exprimés (GDE). Nous avons exploité ces résultats à l'aide de plusieurs outils bio-informatiques. Une analyse **Gene Ontologique** (GO) a permis de classer ces GDE dans 14 processus biologiques. Des tests d'enrichissement ont mis en évidence un enrichissement des GO associés au métabolisme lipidique. Des outils d'enrichissement de voies métaboliques et de construction de réseaux nous ont ensuite permis d'identifier les pathways des acides gras et de leur dégradation en ciblant notamment deux régulateurs clés, la **leptine** et le facteur de transcription **HNF4A**. Nous avons ainsi pu mettre en évidence que ces GDE dans les tissus adipeux de chèvres laitières sont surtout impliqués dans **l'oxydation et la synthèse des lipides**.

Ces résultats ne sont que des premières hypothèses et ces analyses devront être poursuivies en les confrontant aux données zootechniques et aux annotations caprines. Cette **stratégie d'exploitation des données** pourra servir à l'analyse de nouvelles données RNA-seq.

## ABSTRACT

The objective of this study was to compare the **lipid storage mechanisms** in **adipose tissue** of **lactating goats** (in phase of mobilization of fat reserves) and non-lactating (lipid storage phase). A high-throughput RNA sequencing (**RNA-seq**) extracted from adipose tissue showed that 252 genes were differentially expressed. We have exploited these results using several bioinformatics tools. A **Gene Ontological** (GO) analyzes allowed classifying these Differentially Expressed Genes (DEG) in 14 biological processes. Enrichment tests were then performed and revealed an enrichment of GO associated with lipid metabolism. Metabolic pathways and network construction enrichment tools have enabled us to identify the pathways of fatty acid and their degradation, specifically targeting two key regulators, **leptin** and **HNF4A** transcription factor. Thus, we were able to show that these DEG in the adipose tissue of dairy goats are mostly involved in **oxidation** and **lipid synthesis**.

These results are just the first hypotheses and these analyses must be performed by confronting with zootechnic data and caprine annotations. The **operating strategy of the data** could be used for the analysis of new RNA-seq data.

# INTRODUCTION

L'Institut National de la Recherche Agronomique, Inra, est un établissement public dans le domaine scientifique et technologique. Il est aujourd'hui le premier institut de recherche agronomique en Europe et le deuxième dans le monde. Mon stage s'est déroulé dans l'Unité Mixte de Recherche sur les Herbivores (UMRH) dont l'objectif principal est de contribuer au développement de systèmes durables d'élevage des herbivores.

L'équipe de recherche Biomarqueurs (BIOM) étudie le fonctionnement des tissus afin de prédire et d'améliorer la qualité des produits des ruminants et notamment le lait. En effet la filière laitière est un atout important pour l'économie française : elle se situe au 2<sup>ème</sup> rang du secteur agro-alimentaire derrière la filière viande pour ce qui est du chiffre d'affaire et représente 67 400 exploitations laitières et 250 000 emplois sur tout le territoire français<sup>1</sup>. La compréhension des mécanismes régissant la lactation est donc un enjeu majeur pour la recherche agronomique.

Une étude sur la comparaison entre le tissu adipeux de chèvres lactantes et non lactantes a été réalisée et des données transcriptomiques ont été obtenues. Mon travail a consisté à analyser ces données issues de séquençage haut débit afin de comprendre les mécanismes de stockage lipidique dans les adipocytes durant la lactation.

Dans ce rapport nous aborderons tout d'abord la structure au sein de laquelle j'ai effectué mon stage, avant d'expliquer plus en détail les mécanismes du tissu adipeux. Nous verrons ensuite les outils utilisés pour exploiter les données transcriptomiques puis nous terminerons par les résultats obtenus et leur interprétation.

---

<sup>1</sup> Agreste, « Enquête annuelle laitière 2014 ».

# 1. STRUCTURE D'ACCUEIL

## 1.1. L'INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA)

L'Institut National de la Recherche Agronomique (Inra) est un organisme de recherche scientifique public finalisé, placé sous la double tutelle du Ministère de l'Enseignement Supérieur et de la Recherche et du Ministère de l'Agriculture et de la Pêche. Créé en 1946, son premier objectif était la modernisation de l'agriculture française afin d'obtenir une production alimentaire nationale suffisante et répondant aux attentes de la société<sup>2</sup>. Aujourd'hui l'Inra est le 1<sup>er</sup> institut de recherche agronomique en Europe avec 8417 chercheurs, ingénieurs et techniciens et se situe au 2e rang mondial pour ses publications en sciences agronomiques, contribuant ainsi à la production de connaissances et à l'innovation dans l'alimentation, l'agriculture et l'environnement. L'Institut déploie sa stratégie de recherche en mobilisant ses 13 départements scientifiques et en s'appuyant sur un réseau unique en Europe, fort de plus de 200 unités de recherche et de 50 unités expérimentales implantées dans 17 centres en région. Son ambition est, dans une perspective mondiale, de contribuer à assurer une alimentation saine et de qualité ainsi qu'une agriculture compétitive et durable, et ce au sein d'un environnement préservé et valorisé<sup>3</sup>.

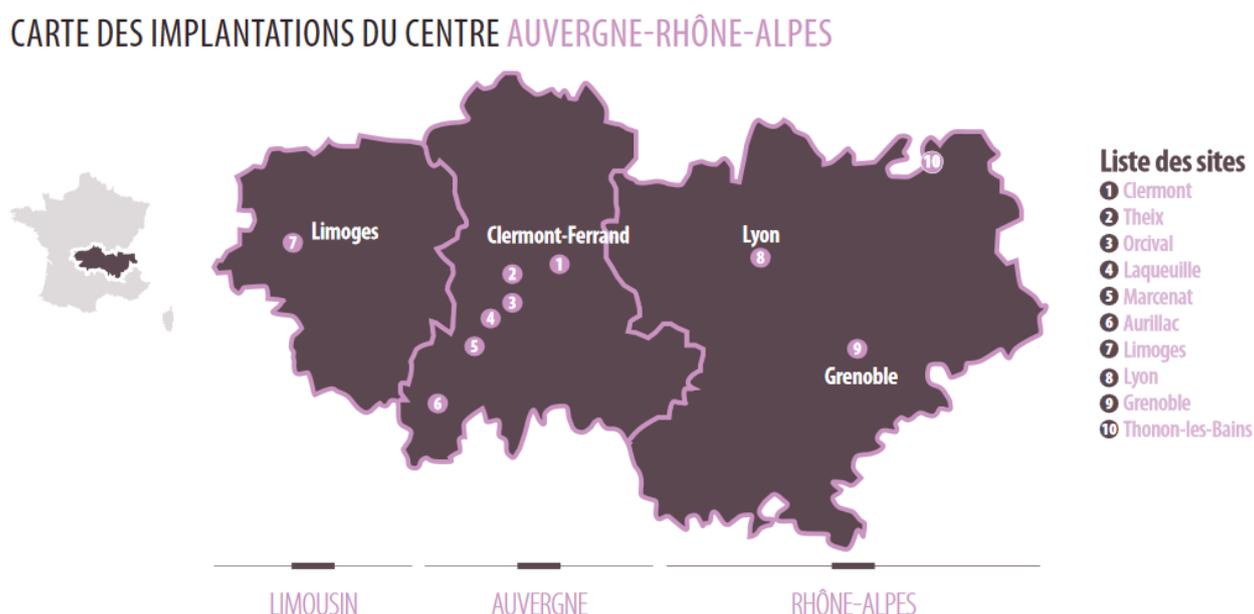
---

<sup>2</sup> INRA, « Jalons historiques ».

<sup>3</sup> INRA, « Les chiffres clés ».

## 1.2. LE CENTRE DE RECHERCHE AUVERGNE-RHÔNE-ALPES

Le centre de recherche Auvergne-Rhône-Alpes (ARA), dans lequel j'ai réalisé mon stage, rassemble des équipes basées en Auvergne, en région Rhône-Alpes mais aussi à Limoges.



**Figure 1** : Carte des implantations du Centre Inra Auvergne-Rhône-Alpes. (Plaquette des activités du centre de recherche Auvergne-Rhône-Alpes, mai 2015).

Ce centre se concentre sur les secteurs de l'agriculture, de l'environnement et de l'alimentation en relation avec la demande sociétale. Ces recherches multidisciplinaires portent sur le développement d'exploitations agricoles durables et leur insertion dans le territoire, sur l'utilisation de la prairie par les herbivores, sur la qualité des produits animaux (viande, lait, fromages) et sur la nutrition humaine. La recherche sur les végétaux est quant à elle abordée par des approches centrées autour de l'adaptation des céréales et de l'arbre à l'environnement<sup>4</sup>.

---

<sup>4</sup> INRA, « Présentation du Centre ».

Fort de 935 agents, dont 761 titulaires (dont environ 40% de scientifiques et d'ingénieurs), répartis dans 31 unités, le centre ARA constitue environ 10% du potentiel de recherche national de l'Inra. Il est le 3ème des 17 Centres de recherches de l'Inra<sup>5</sup>.

### 1.3. L'UNITE MIXTE DE RECHERCHE SUR LES HERBIVORES (UMRH)

Créée en 2012, l'UMR1213 Herbivores est une unité mixte entre l'Inra et l'Institut d'enseignement supérieur et de recherche en alimentation, santé animale, sciences agronomiques et de l'environnement (VetAgro Sup). Cette unité contribue au développement de systèmes durables d'élevage des herbivores, conciliant efficacité des productions, qualité des produits et viabilité socio-économique avec protection et valorisation de l'environnement, et bien-être animal. Nécessitant des approches analytiques fines, comme les méthodes à haut débit dites Omiques, les recherches engagées par l'UMRH visent à identifier et maîtriser les déterminants des systèmes étudiés (exploitation, troupeau, animal, fonctions biologiques). Elles mobilisent les moyens des unités expérimentales Inra de l'Herbipôle et de la Plate-Forme d'Exploration du Métabolisme de Theix. Les résultats permettent une meilleure compréhension des mécanismes biologiques qui sous-tendent les phénotypes et aboutissent à la proposition d'outils d'aide à la décision à destination des différents acteurs de la société (producteurs, consommateurs, citoyens et décideurs publics). L'UMRH est constituée de 133 agents titulaires dont 74 chercheurs et ingénieurs ainsi qu'une trentaine d'agents non titulaires par an dont une trentaine de doctorants et post-doctorants.

---

<sup>5</sup> Auvergne Sciences, « Annuaire - INRA - Institut National de la Recherche Agronomique ».

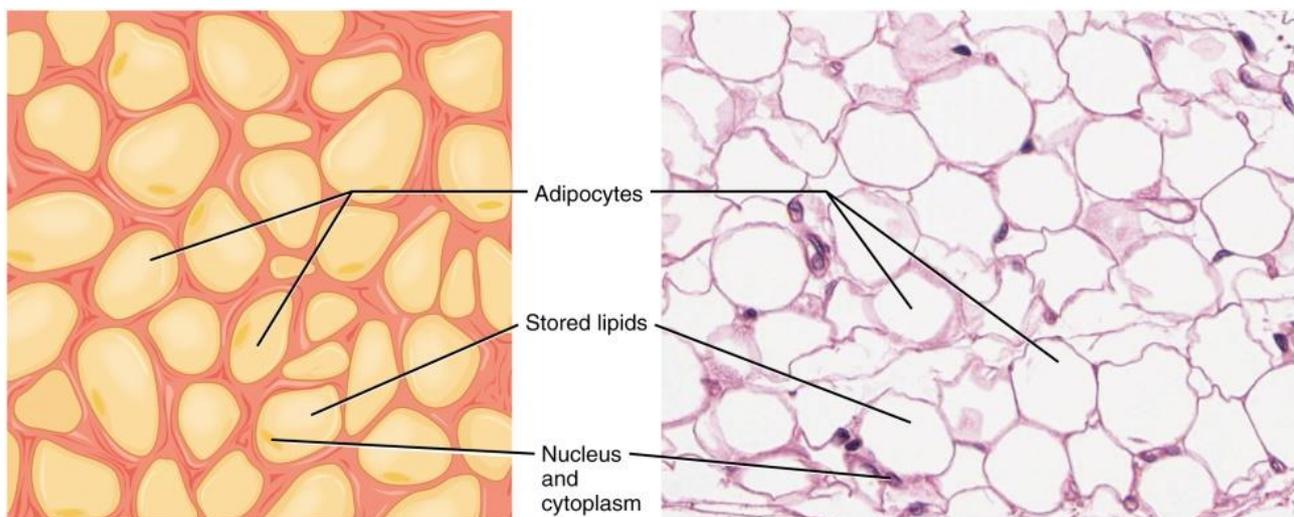
#### 1.4. L'EQUIPE BIOMARQUEURS (BIOM)

L'équipe Biomarqueurs étudie le fonctionnement des tissus dans le but de prédire les performances, l'adaptation et la qualité des produits de ruminants (viande et lait). Cette équipe a notamment pour objectif de déterminer les mécanismes biochimiques et moléculaires du fonctionnement des tissus (muscle, glande mammaire...) ainsi que d'analyser les réponses de ces tissus et de l'animal aux facteurs d'élevage tel que l'alimentation et son interaction avec le génotype, le statut physiologique et/ou l'état sanitaire de l'animal. Une autre partie des activités de l'équipe BIOM est la conception d'outils de phénotypage de ces caractères de production, basés sur la détection et la mesure de marqueurs moléculaires.

## 2. ETUDE BIBLIOGRAPHIQUE DU SUJET

### 2.1. GENERALITES SUR LE TISSU ADIPEUX (TA)

Le tissu adipeux (TA) est un tissu conjonctif, ses cellules étant séparées par une matrice extracellulaire. Il est constitué de cellules grasses appelées « adipocytes » (Figure 2) qui ont la particularité de posséder un réservoir lipidique important. Le rôle du TA est d'offrir une importante réserve de graisses (lipides) qui peut être mobilisée selon les besoins de l'organisme. Le métabolisme du TA peut se schématiser en 2 parties : la lipogénèse et la lipolyse.

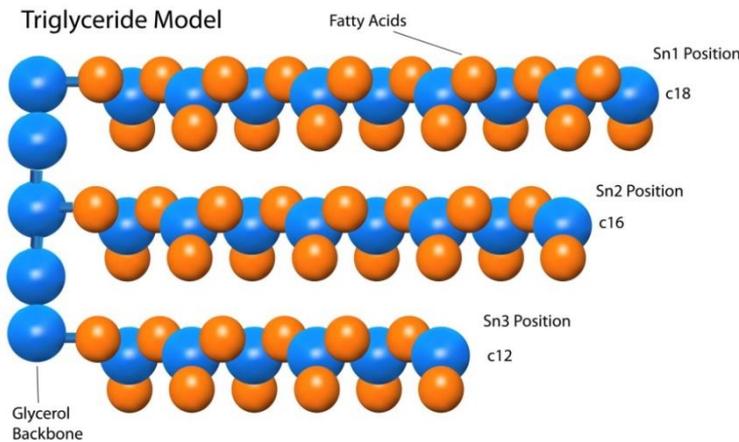


**Figure 2:** Représentation schématique et photo au microscope de cellules adipeuses (Anatomy & Physiology par OpenStax, chapitre 4.3, 2013).

### 2.1.1. LIPOGENESE

La lipogenèse est la biosynthèse d'acides gras (AG) à partir de précurseurs qui conduit à la formation de triglycérides (TG) composés de 3 AG reliés à une molécule de glycérol par des liaisons esters (Figure 3). Elle permet la mise en réserve sous forme de TG de l'énergie excédentaire apportée par l'alimentation. La lipogenèse regroupe les activités cellulaires<sup>6</sup> de :

- Synthèse *de novo* d'AG à partir de divers précurseurs (principalement l'acétate, le glucose et le lactate chez le ruminant). Ce mécanisme est notamment régulé par la concentration en glucose (qui le stimule) et la concentration en AG long (qui l'inhibe)<sup>7</sup>.
- Captation des AG des TG circulants d'origine alimentaire après action de la lipoprotéine-lipase (LPL) dans le sang. En effet les AG circulent dans le sang sous formes de TG incorporés dans des lipoprotéines. Afin qu'ils franchissent les barrières membranaires il est nécessaire de les découper grâce à la LPL. Cette enzyme n'agit pas sur les TG stockés dans la cellule adipeuse et elle est régulée entre autre par l'insuline (activation) et les AG longs (inhibition).
- Puis d'estérification de ces AG en TG stockés dans la gouttelette lipidique de l'adipocyte.



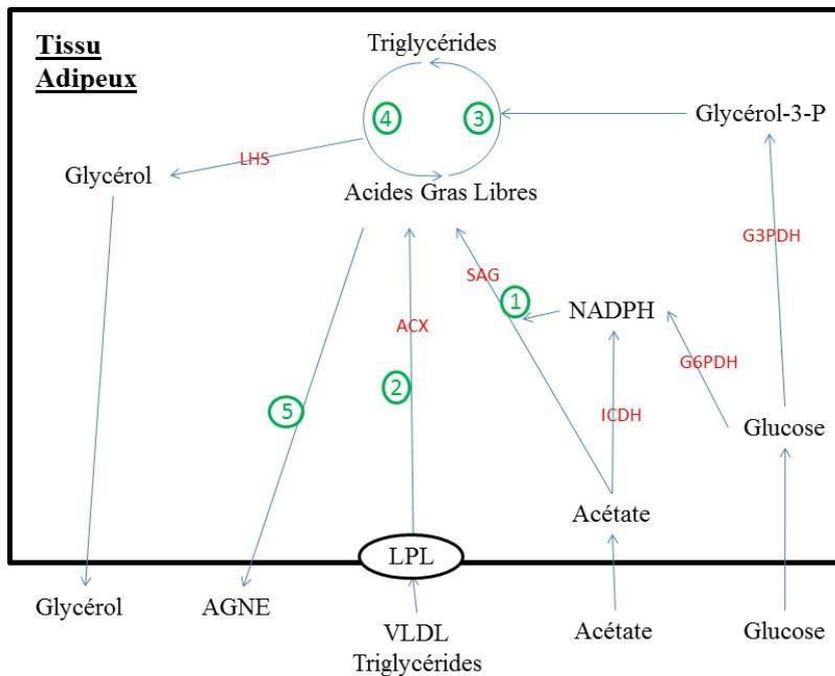
**Figure 3 :** Molécule de Triglycéride (Roger DANIELS, 2010).

<sup>6</sup> Chilliard, « Métabolisme du tissu adipeux, lipogénèse mammaire et activités lipoprotéine-lipasiqes chez la chèvre au cours du cycle gestation-lactation ».

<sup>7</sup> Faulconnier, Delavaud, et Chilliard, « Effets de l'insuline et/ou de la dexaméthasone sur les activités enzymatiques et métaboliques des tissus adipeux bovin et ovin maintenus en culture durant 7 jours ».

### 2.1.2. LIPOLYSE

La lipolyse regroupe les activités cellulaires d'hydrolyse des liaisons esters entre le glycérol et les AG, sous l'effet de la lipase hormono-sensible (LHS) et de l'adipose TG lipase, puis de libération d'une partie de ceux-ci dans le sang. Les AG libérés s'associent dans la circulation sanguine avec l'albumine sérique. Le glucose et l'insuline, entre autres, inhibent la lipolyse alors que l'adrénaline et le glucagon exercent sur cette voie un rétrocontrôle positif<sup>8</sup> (Figure 4).



Principales voies du métabolisme lipidique dans le tissu adipeux de ruminant :

- |  |  |
|--|--|
| 1 - Lipogénèse de novo                         | ACX : Acétyl CoA Carboxylase               |
| 2 - Hydrolyse et prélèvement des triglycérides | GPDH : Glycérol-3-Phosphate Déshydrogénase |
| 3 - (ré) estérification des acides gras        | ICDH : Isocitrate Déshydrogénase           |
| 4 - Lipolyse                                   | LPL : Lipoprotéine Lipase                  |
| 5 - Lipomobilisation                           | LHS : Lipase Hormono-Sensible              |
|  | SAG : Synthétase des Acides Gras (ou FAS)  |
|  | VLDL : Lipoprotéine de très faible densité |

**Figure 4 :** Schéma récapitulatif du métabolisme des lipides dans le tissu adipeux de ruminant (Y. Chilliard, 1985).

<sup>8</sup> Bonnet et Chilliard, « Régulation par l'état nutritionnel et la photopériode de l'expression de gènes spécifiant des enzymes du métabolisme énergétique dans le tissu adipeux et le muscle de brebis. »

### 2.1.3. CYCLES LIPOGENESE/LIPOLYSE

Au cours des cycles de production, les ruminants femelles sont soumis pour des raisons physiologiques, économiques ou saisonnières à des périodes de sous-nutrition de durée et d'intensité variables. Les TA sont alors mobilisés et doivent être ensuite reconstitués efficacement. Inversement, on peut rechercher à limiter le dépôt de TA chez des animaux producteurs de viande, pour des raisons économiques et/ou diététiques. La succession des cycles de lipogénèse-lipolyse permet donc à l'animal de s'adapter (en les anticipant) aux fluctuations des besoins physiologiques (notamment en fin de gestation et début de lactation) et des apports nutritionnels (disponibilités saisonnières en fourrages, limitations physiologiques de l'appétit). Chez les ruminants, le TA est le principal site de synthèse d'AG (sauf pendant la lactation où la mamelle joue un rôle majeur). Chez d'autres espèces (rongeurs, homme, oiseaux...), le foie joue un rôle primordial<sup>9</sup>.

## 2.2. METABOLISME DU TISSU ADIPEUX CHEZ LE RUMINANT EN LACTATION.

En début de lactation, la prise alimentaire est insuffisante pour subvenir aux besoins en nutriments de la glande mammaire. Les animaux mobilisent donc leurs réserves stockées précédemment dans le TA sous forme de TG. Cette mobilisation des réserves adipeuses se reflète par une diminution de la lipogénèse au profit de la lipolyse comme c'est le cas en fin de gestation. Les réserves lipidiques contenues dans le TA vont être utilisées à la fois par les glandes mammaires et par les autres tissus maternels. Cette mobilisation intensive des réserves adipeuses se traduira par une diminution du volume des adipocytes. Il est à noter qu'en fin de lactation l'animal arrive généralement à équilibrer sa balance énergétique. En effet les tissus maternels s'adaptent aux nouvelles exigences des glandes mammaires en termes de glucides, lipides et acides aminés<sup>10</sup>.

---

<sup>9</sup> Faulconnier, Delavaud, et Chilliard, « Effets de l'insuline et/ou de la dexaméthasone sur les activités enzymatiques et métaboliques des tissus adipeux bovin et ovin maintenus en culture durant 7 jours ».

<sup>10</sup> Chilliard, « Métabolisme du tissu adipeux, lipogénèse mammaire et activités lipoprotéine-lipasiques chez la chèvre au cours du cycle gestation-lactation ».

Des mécanismes de régulation transcriptionnelle et post-transcriptionnelle interviennent durant la lactation dans le TA de ruminant, modifiant l'activité de plusieurs enzymes lipogéniques. Par exemple, en début de lactation où la balance énergétique est négative, une diminution de l'expression des gènes clés de la lipogénèse comme la LPL, impliquée dans la captation des AG, ou la synthétase des AG et l'enzyme malique, toutes deux impliquées dans la synthèse *de novo* d'AG, est observée alors que l'expression des enzymes clés de la lipolyse (LHS) est augmentée<sup>11</sup>.

Ces adaptations métaboliques du TA durant la lactation sont également contrôlées par plusieurs systèmes endocrines et neurocrines, faisant intervenir notamment l'hormone de croissance et la prolactine qui sont augmentées alors que l'insuline (INS), la leptine et les hormones thyroïdiennes sont diminués<sup>12</sup>.

---

<sup>11</sup> Faulconnier et al., « The Transcriptomic Profiles of Adipose Tissues Are Modified by Feed Deprivation in Lactating Goats »; Sumner-Thomson, Vierck, et McNamara, « Differential expression of genes in adipose tissue of first-lactation dairy cattle ».

<sup>12</sup> Vernon, « Lipid metabolism during lactation »; Khan et al., « Leptin as a Modulator of Neuroendocrine Function in Humans ».

## 3. MATERIEL ET METHODE

### 3.1. LE PROJET MILKCHEST

Ce stage s'inscrit dans un projet ANR (Agence Nationale de la Recherche) intitulé "Approche intégrée des voies de sécrétion des protéines et des lipides dans la cellule épithéliale mammaire : dissection moléculaire de l'adaptabilité de la matière grasse laitière" (MilkChEST) dont l'objectif principal est d'évaluer et de décrypter les mécanismes de biosynthèse et de sécrétion des principaux constituants du lait chez la chèvre laitière<sup>13</sup>. Afin de mieux préciser ces voies de sécrétion, leur comparaison a été réalisée avec une situation de stockage des lipides obtenue dans le TA. Pour ce faire, et dans le cadre de mon stage, le niveau d'expression de l'ensemble des gènes (appelé transcriptome) provenant de TA de chèvres lactantes (cellules considérées comme étant en phase de non stockage lipidique) a été comparé à celui du TA de chèvres en période sèche (non lactantes, cellules considérées comme étant en phase de stockage lipidique).

La caractérisation et la quantification du transcriptome dans un tissu donné et dans des conditions données permettent d'identifier sans à priori l'ensemble des gènes exprimés, et ainsi de déterminer les mécanismes de régulation et les réseaux d'expression des gènes. Les techniques couramment utilisées pour étudier le transcriptome incluent les puces à ADN et, plus récemment, le séquençage d'ARN à haut débit dit RNA-Seq ou NGS pour *Next Generation Sequencing*.

Ma mission durant mon stage a été de réaliser une analyse bio-informatique des données d'expression génique obtenues par séquençage haut débit des ARN extraits des TA de chèvres lactantes et de chèvres non lactantes, pour en dégager les processus biologiques, les voies métaboliques et les réseaux de régulations les plus discriminants dans ces deux conditions physiologiques.

---

<sup>13</sup> Leroux et al., « MilkChEST, an integrative approach of the secretion pathways in the mammary epithelial cell ».

### 3.2. ETUDE EXPERIMENTALE

Afin d'étudier les mécanismes de stockage lipidique chez la chèvre, 4 chèvres non lactantes et 4 chèvres lactantes (au 74<sup>ème</sup> jour de lactation) de race Alpine ont été sacrifiées et un site de TA interne (TA omental) prélevé afin d'obtenir des échantillons de TA, a priori respectivement en situation de stockage ou non. Elles ont toutes reçu le même régime alimentaire composé de foin de luzerne distribué à volonté, complémenté avec du concentré. Les extractions d'ARN totaux ont été réalisées avant mon arrivée au laboratoire sur le TA de ces 8 chèvres, en utilisant le kit Invitrogen, et ont montré une qualité d'ARN excellente avec un RIN (*Rna Integrity Number*) moyen égal à 9 pour les analyses ultérieures (Tableau I). Un séquençage haut débit de ces ARNm a ensuite été réalisé par la plateforme de séquençage « Genomic Paris Centre » début 2016.

**Tableau I:** Concentration et qualité des ARNm extraits des tissus adipeux omental des chèvres lactantes et non lactantes. (Source personnelle).

N° Animal	N° Sample pour Sequençage	Etat Physiologique	Concentration (ng/µl)	RIN
13263	Sample82	Lactante	23,895	8,8
13060	Sample83	Lactante	25,525	8,3
13150	Sample84	Lactante	24,235	8,7
13566	Sample85	Lactante	26,8	9,5
11053	Sample86	Non Lactante	36,55	8,7
11315	Sample87	Non Lactante	29,61	8,9
90179	Sample88	Non Lactante	35,38	8,7
90222	Sample89	Non Lactante	31,83	8,9

### 3.3. LE SEQUENÇAGE ILLUMINA

Afin d'analyser les séquences d'ARN extraites de nos tissus, la plateforme a procédé à une analyse de données RNA-Seq aussi appelé "séquençage de seconde génération" (*next-generation sequencing*) ou "séquençage à haut débit" (*high-throughput sequencing*) à l'aide d'un séquenceur de type Illumina.

Cette analyse se divise en 2 étapes :

- préparation et amplification des molécules d'ARN à analyser : après fragmentation et ancrage des ARN sur un support solide, ceux-ci subissent une étape d'amplification par PCR par pont (formation de clusters). (Figure 5)
- incorporation des bases complémentaires du brin à séquencer et lecture de la séquence : la réaction de séquençage se déroule position après position en ajoutant un mélange contenant toutes les bases associées chacune à un fluorochrome différent. L'extrémité de ces bases est protégée pour empêcher l'addition de bases supplémentaires à chaque cycle d'incorporation. Une lecture laser permet alors de détecter simultanément toutes les positions incorporées. Le clivage des fluorochromes permet ensuite l'incorporation de la base suivante. La lecture est effectuée ainsi cycle après cycle (Figure 6). A chaque base enregistrée est associé un score de qualité qui peut diminuer du fait de l'intégration de bases différentes au sein d'un même cluster<sup>14</sup>.

Cette méthode permet l'acquisition en parallèle de plus de 3 milliards de séquences de 100 bases de long, appelées *reads*. Dans ce projet, le but de cette technique est d'estimer l'abondance de chaque *read* et ainsi l'abondance des ARN correspondants dans l'échantillon étudié<sup>15</sup>.

La plateforme Genomic Paris Centre s'est ensuite chargée de traiter et d'analyser les données issues du séquençage :

- filtrage des séquences brutes notamment en fonction des scores de qualité du séquençage et de la taille des *reads*.

alignement des séquences contre le génome de référence : notre sujet d'étude étant la chèvre, il aurait été souhaitable d'utiliser des bases de données *Capra hircus*.

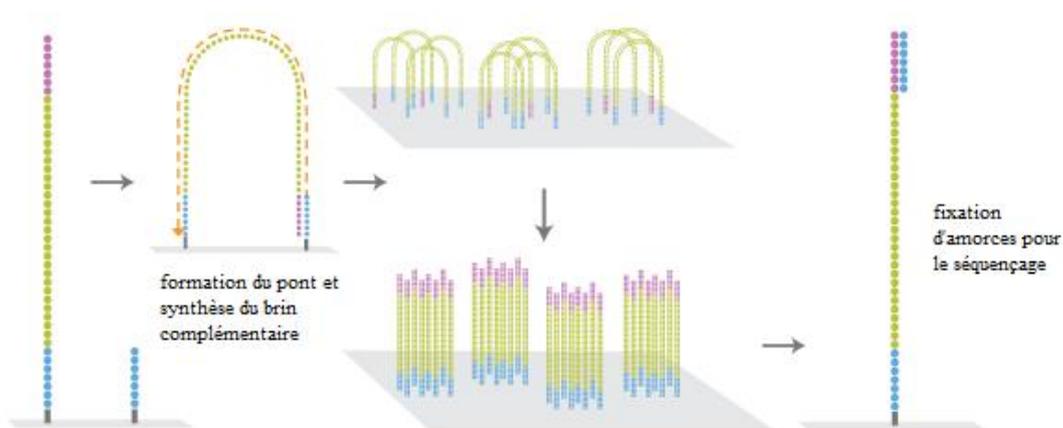
---

<sup>14</sup> J. Delafontaine, « L'analyse de données RNA-seq »; Oshlack, Robinson, et Young, « From RNA-Seq Reads to Differential Expression Results ».

<sup>15</sup> Le Crom et al., « Concept paper : Le séquençage a haut débit méthodes et enjeux en médecine, pharmacologie et toxicologie ».

Cependant, la séquence complète du génome caprin n'étant pas totalement annotée à l'heure actuelle, contrairement à celui de l'ovine, le génome bovin (*Bos taurus*) a été utilisé comme génome de référence pour l'exploitation des résultats. Des études de génomique structurale ont néanmoins montré que ces deux espèces étaient très proches<sup>16</sup>.

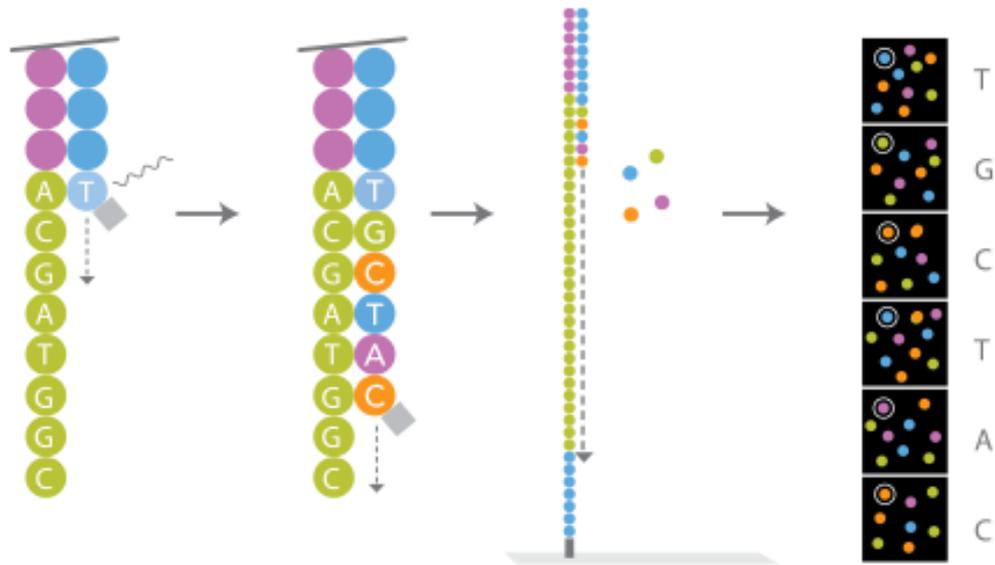
- nouvelle étape de filtrage afin d'éliminer les *reads* ne s'alignant pas et ceux s'alignant à plusieurs endroits avec une affinité similaire.
- comptage des lectures par rapport à l'annotation du génome déjà existant, suivi d'une normalisation des comptages afin d'éviter les biais dû à la différence de profondeur de séquençage. (nombre moyen de *reads* par position alignée). En effet les échantillons avec une plus grande profondeur de séquençage et donc plus de *reads* seront grandement favorisés lorsqu'il faudra les comparer. La normalisation des données a été faite grâce à la librairie DESeq du logiciel de statistique R.



**Figure 5 :** Phase d'amplification par PCR par pont durant un séquençage Illumina. (The Illumina HiSeq 2000 Sequencing Technology, German Cancer Research Center, 2015).

---

<sup>16</sup> Schibler et al., « Comparative Gene Mapping ».



**Figure 6** : Phase de séquençage par excitation laser des fluorochromes durant un séquençage Illumina.

(The Illumina HiSeq 2000 Sequencing Technology, German Cancer Research Center, 2015).

Les résultats de ces analyses ont été fournis sous la forme d'un tableau avec comme principales informations :

- Les ENSEMBLEID
- Le nombre moyen de comptage par chèvre (chaque échantillon ayant été séquencé 4 fois)
- Le nombre moyen de comptage pour les chèvres lactantes et non lactantes
- Les *FoldChanges* et les *log2FoldChanges* (comptage lactantes / non lactantes)
- Les *pvalues* et les *pvalues* ajustées avec DESeq
- Les *GENE SYMBOL*
- Les identifiants HGNC (HUGO *Gene Nomenclature Committee*)
- Les emplacements dans le génome et sur le chromosome

### 3.4. ANALYSE BIO-INFORMATIQUE DES DONNEES

J'ai utilisé au cours de cette analyse divers outils bio-informatiques afin d'identifier le rôle des gènes différentiellement exprimés (GDE) dans nos deux conditions. Ces applications n'utilisant pas les mêmes bases de données, les mêmes algorithmes, ni les mêmes manières de procéder, il est important d'utiliser un large éventail d'outils afin d'obtenir un maximum de résultats pour ensuite les croiser et en dégager, si possible, les processus biologiques, les voies métaboliques et les réseaux de régulations les plus impactés dans nos conditions expérimentales et ainsi pouvoir au mieux interpréter nos résultats

#### 3.4.1. *PROTEINSIDE*

ProteINSIDE est le premier outil gratuit en ligne, développé par l'Inra, pour la fouille des données de génomique et de protéomique chez le bovin, l'ovin, et le caprin<sup>17</sup>. Il intègre également des données homme, rat et souris. Cet outil 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 (*Gene Ontology*, GO)<sup>18</sup>. La GO consiste en un vocabulaire contrôlé et structuré hiérarchiquement permettant de décrire les fonctions des produits de gènes selon trois axes : les processus biologiques (BP), les composants cellulaires (CC) et les fonctions moléculaires (MF) associées. Il prédit également 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. L'interface web permet la création d'analyses et la visualisation des résultats en temps réel ou leurs téléchargements dans de nombreux formats (images, PDF, tableaux, données brutes, fasta...).

---

<sup>17</sup> Kaspric et al., « ProteINSIDE to Easily Investigate Proteomics Data from Ruminants ».

<sup>18</sup> Ashburner et al., « Gene Ontology ».

### 3.4.2. *PROTEIN ANALYSIS THROUGH EVOLUTIONARY RELATIONSHIPS (PANTHER)*

Panther est une base de données gratuite qui permet d'identifier et de classer les gènes et leurs protéines en s'appuyant sur les données GO.

PANTHER dispose de plusieurs outils :

- Une fonction *Gene List Analysis* permettant d'effectuer une analyse fonctionnelle à partir d'une liste des gènes. Panther va alors tracer un diagramme circulaire et hiérarchique des GO correspondants. En cliquant sur les quartiers, il est possible de voir les GO plus spécifiques (Figure 7).
- Une fonction d'analyse d'enrichissement c'est-à-dire une recherche des termes GO surexprimés significativement dans un jeu de données. Cet enrichissement des termes peut permettre de définir des caractéristiques fonctionnelles possibles du jeu de données. Une approche statistique est utilisée, associant à chaque terme enrichi une valeur de significativité (*p-value*). Les résultats sont présentés sous la forme d'un tableau hiérarchique.

## PANTHER GO-Slim Biological Process

Total # Genes: 237 Total # process hits: 487



- [apoptotic process \(GO:0006915\)](#)
- [biological adhesion \(GO:0022610\)](#)
- [biological regulation \(GO:0065007\)](#)
- [cellular component organization or biogenesis \(GO:0071840\)](#)
- [cellular process \(GO:0009987\)](#)
- [developmental process \(GO:0032502\)](#)
- [growth \(GO:0040007\)](#)
- [immune system process \(GO:0002376\)](#)
- [localization \(GO:0051179\)](#)
- [locomotion \(GO:0040011\)](#)
- [metabolic process \(GO:0008152\)](#)
- [multicellular organismal process \(GO:0032501\)](#)
- [reproduction \(GO:0000003\)](#)
- [response to stimulus \(GO:0050896\)](#)

**Figure 7:** Répartition des gènes différentiellement exprimés selon leurs ontologies obtenues avec le logiciel PANTHER (Source personnelle d'après Panther et VisiBone).

### 3.4.3. KEGG (KYOTO ENCYCLOPEDIA OF GENES AND GENOMES)

KEGG est un outil gratuit qui intègre à la fois des informations génomiques, chimiques et fonctionnelles. Il est constitué de 15 bases de données principales et permet notamment la représentation graphique de systèmes moléculaires.

KEGG MAPPER est l'interface utilisateur de la cartographie de KEGG. *Search&color Pathway* est un outil de KEGG MAPPER qui permet d'identifier les voies cellulaires dans lesquelles sont impliquées nos gènes d'intérêts<sup>19</sup>. Ainsi, il est possible de visualiser l'ensemble d'une voie cellulaire donnée avec les interactions entre gènes impliquées. Cet outil permet de noter par des codes couleurs les gènes afin de les différencier en fonction de leur expression (bleu pour les sous-exprimés, rouge pour les surexprimés).

### 3.4.4. PATHWAY STUDIO

Pathway Studio est un logiciel payant développé pour la visualisation et l'analyse des voies métaboliques (pathway), des réseaux de régulation des gènes et des interactions entre protéines. Pour établir ces relations il réalise du *text mining*<sup>20</sup> qui consiste à extraire des informations d'articles scientifiques et à les interpréter pour relier des éléments (gènes, protéines...) entre eux de manière précise en définissant le type de relation et en mettant à disposition des liens vers les articles traitant du sujet<sup>21</sup>. La base de données ResNet, utilisée par Pathway Studio, contient 3.7 millions d'articles provenant de 1496 journaux scientifiques, 30 millions de phrases, 4.7 millions de relations extraites de ces phrases et 1300 voies métaboliques scientifiquement validées<sup>22</sup>. En entrant un jeu de donnée avec les ratios d'expression des gènes, Pathway Studio est aussi capable d'effectuer des analyses d'enrichissements de pathways ou de GO.

---

<sup>19</sup> Kanehisa et al., « KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets ».

<sup>20</sup> Jenssen et al., « A Literature Network of Human Genes for High-Throughput Analysis of Gene Expression ».

<sup>21</sup> Bonnet et al., « Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares »; Nikitin et al., « Pathway Studio—the Analysis and Navigation of Molecular Networks ».

<sup>22</sup> « Pathway Studio - Fact Sheet ».

## 4. RESULTATS ET DISCUSSION

### 4.1. IDENTIFICATION DES GENES

L'exploitation des données de séquençage haut débit a permis de quantifier l'expression de 19 271 gènes dans le TA de chèvres lactantes et non lactantes. La lactation (comparaison des TA des chèvres lactantes contre non lactantes) a entraîné une modification significative ( $P < 0.05$ ) de l'expression de 252 gènes dont 118 surexprimés et 134 sous exprimés. Parmi eux, on dénombre 6 gènes inconnus. Dans cette liste de GDE, nous avons relevé les gènes ayant un nombre de transcrits particulièrement important (moyenne des comptages lactant et non-lactant  $> 30\,000$ ) (Tableau II). Nous avons cherché les orthologues humains de nos gènes alignés sur *Bos Taurus* grâce au module ProteConvert du logiciel Proteinside afin de pouvoir utiliser les EnsemblID contre des bases de données humaines qui sont plus fournies du fait du nombre plus important d'études et d'annotations réalisées sur le génome humain.

La différence d'expression des gènes se mesure grâce au *foldchange* (FC), soit le ratio entre nombre moyen de comptage chez la chèvre lactante par le nombre moyen de comptage chez la chèvre non lactante pour un gène donné. Un résultat supérieur à 1 montre une surexpression du gène chez la chèvre lactante alors qu'un résultat inférieur à 1 montre une surexpression chez les chèvres non lactantes. Dans la plupart des cas le  $\log_2(FC)$  est utilisé afin que les valeurs *UP* ( $> 0$ ) et *DOWN* ( $< 0$ ) régulées soient centrées sur 0.

Pour la suite de mes travaux nous avons utilisé la liste des GDE.

**Tableau II :** Gènes les plus exprimés parmi les gènes différentiellement exprimés (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle).

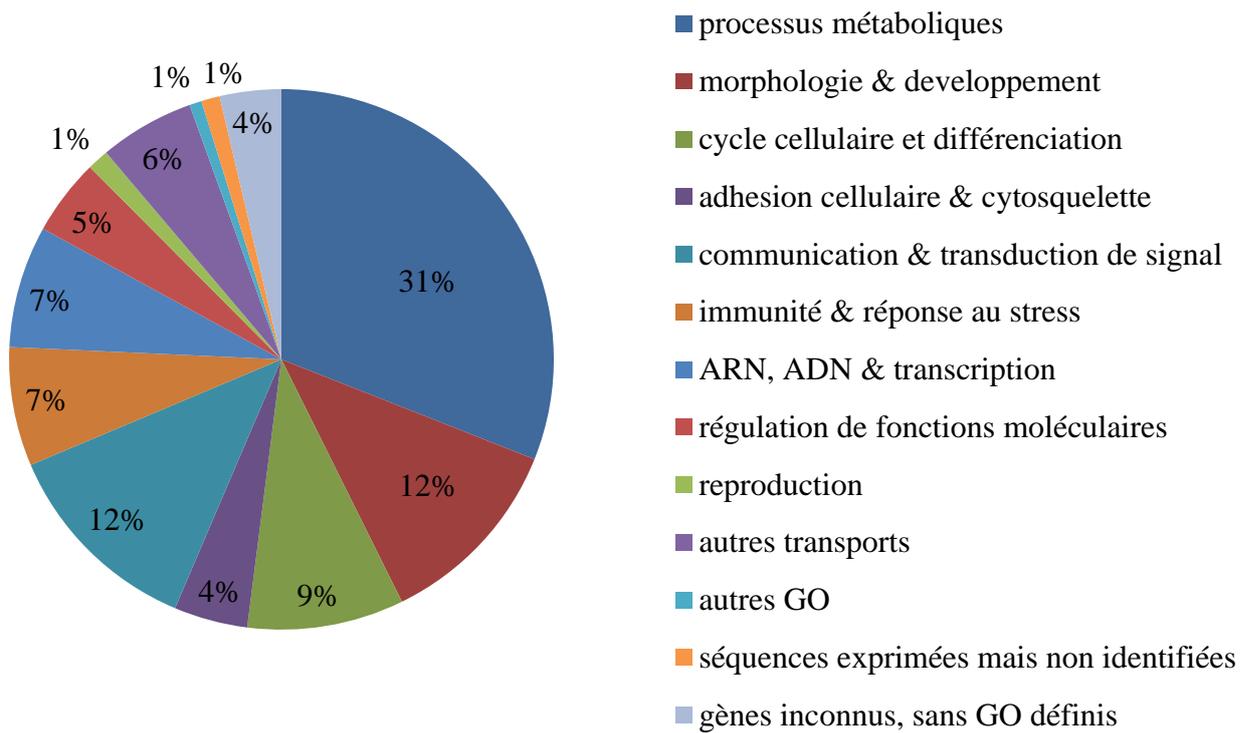
gènes les plus exprimés				
Gene Name	baseMean non-lactar	baseMean lactan	Moyenne comptage	log2FoldChange
AHNAK	166212	81243	247455	-1,03
LPL	61735	153062	214797	1,31
ACSL1	40194	103471	143664	1,36
TLN1	78255	40613	118868	-0,95
PLIN4	73035	33988	107023	-1,10
ANXA8	61414	15879	77293	-1,95
TNS1	50307	24617	74923	-1,03
LGALS1	54304	17904	72208	-1,60
ADIRF	46574	24399	70973	-0,93
ACTN4	45726	21009	66734	-1,12
ALPK3	16299	50732	67031	1,66
GALNT16	41903	20640	62542	-1,02
AQP7	13670	44254	57923	1,69
LEP	48031	9194	57225	-2,39
TIMP3	36274	16599	52873	-1,13
ACAT1	16960	34359	51319	1,02
PCK2	7622	39105	46726	2,36
CHI3L1	28748	13393	42141	-1,10
BOLA	27503	14384	41887	-0,94
ECM2	25694	14469	40163	-0,83
RPS2	13404	25310	38714	0,92
SFMBT2	25815	12062	37877	-1,10
ENSBTAG00000043570	10013	24572	34585	1,30
AKAP12	24381	9977	34358	-1,29
NNAT	5371	27863	33234	2,38
S100A10	19693	10647	30340	-0,89
VCL	20206	10088	30293	-1,00

## 4.2. ONTOLOGIE DES GENES

### 4.2.1. CLASSIFICATION DES GENES PAR GO

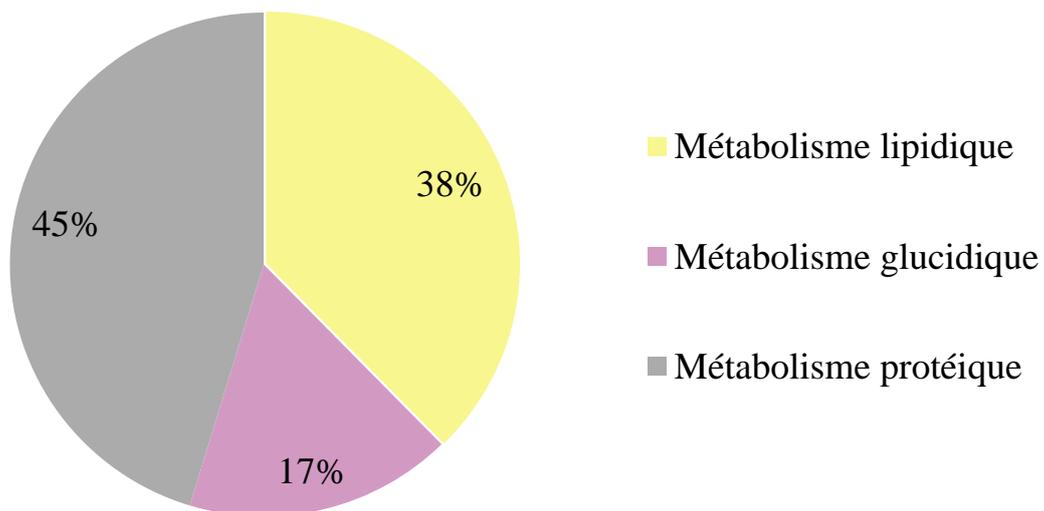
Comme nous l'avons vu précédemment, l'analyse GO a pour but d'identifier et de classer les produits des gènes. Le but de notre expérience étant de comparer les métabolismes lipidiques entre les chèvres lactantes et non lactantes, l'analyse s'est concentrée sur l'étude des BP.

Nous avons utilisé PANTHER qui offre la possibilité de classer les GO de manière hiérarchique sous la forme d'un diagramme circulaire. Il nous a alors été possible de trier les GDE dans 14 processus biologiques différents. Cependant de nombreux GDE n'ont pas été classés par cette méthode et la hiérarchie des GO obtenue ne nous semblait pas totalement pertinente. Par conséquent, nous les avons retriés manuellement en s'aidant de la « *GO database BP Complete* » de PantherDB et en définissant des GO plus parlants pour le chercheur (Figure 8 et Annexe 1). Nous remarquons que 57% des GDE sont classés dans 3 GO : processus métabolique, communication & signal transduction, et morphologie & développement. Notre étude portant sur le TA, nous nous sommes donc focalisés sur le processus métabolique, qui contient 31% de l'ensemble des GDE. En observant les sous-classes de ce GO (Figure 9 et Annexe 2), nous mettons en évidence une prépondérance des métabolismes protéique et lipidique qui représentent plus de 80% des gènes impliqués dans le processus métabolique.



**Figure 8 :** Répartition des gènes différentiellement exprimés selon leurs ontologies après analyse personnelle (Source personnelle).

### Répartition des gènes dans la GO Primary Process



**Figure 9 :** Pourcentage de gènes différentiellement exprimés dans les processus métabolismes lipidique, glucidique et protéique (Source personnelle).

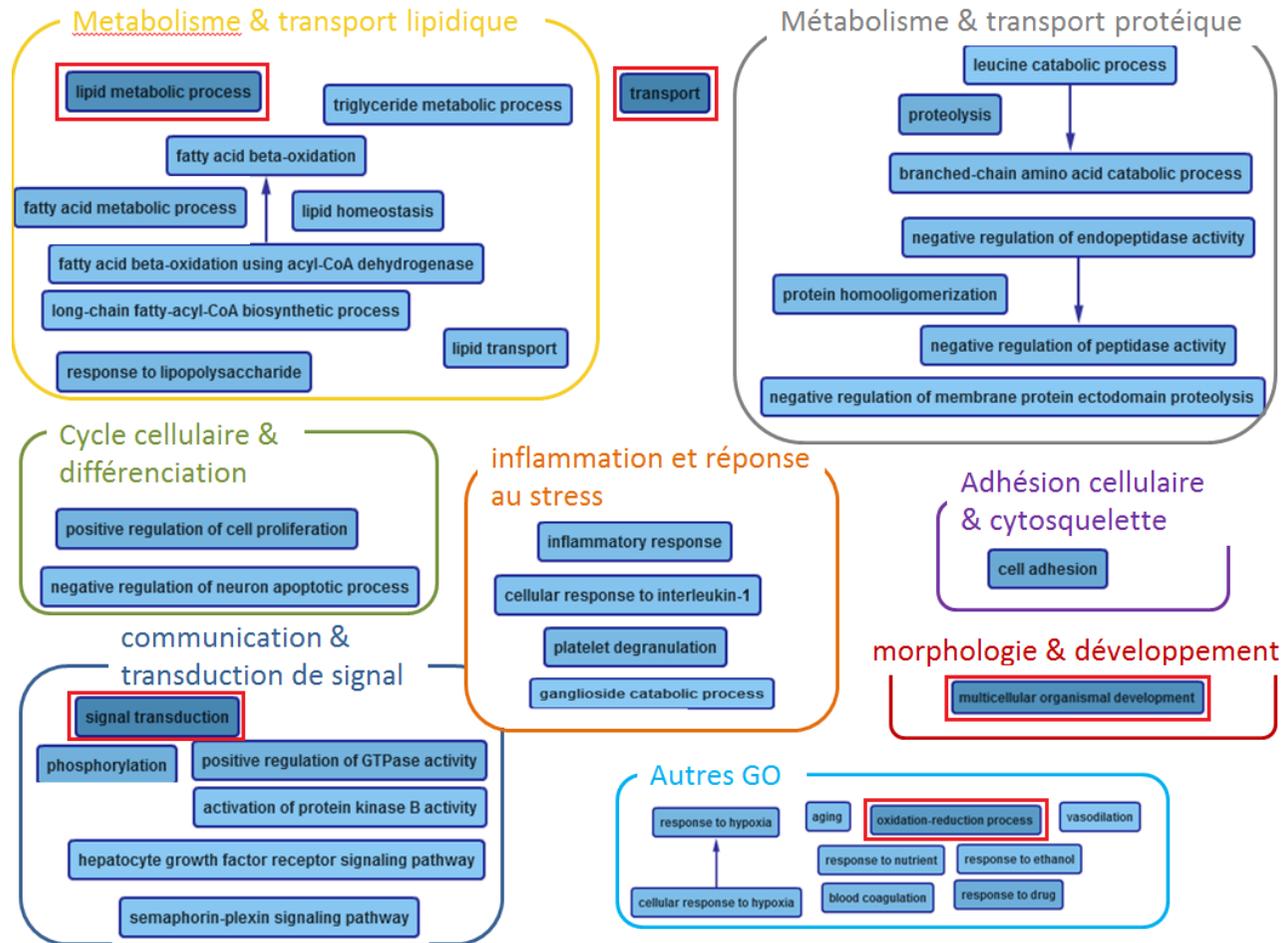
Au vu de ces résultats, il semble qu'il y ait une prédominance de certains groupes ontologiques dans notre jeu de données, en particulier les métabolismes lipidique et protéique, mais nous ne pouvons pas affirmer qu'ils soient enrichis par rapport à un échantillon pris aléatoirement.

#### 4.2.2. ENRICHISSEMENT EN GO

Afin de déterminer les GO significativement surreprésentés dans notre jeu de données nous avons effectué des tests statistiques d'enrichissement.

Les résultats obtenus avec ProteINSIDE mettent en évidence que deux processus biologiques sont particulièrement enrichis : le métabolisme lipidique et la communication & transduction de signaux (Figure 10, les GO les plus enrichies sont encadrées en rouge). Un 2ème test d'enrichissement avec le logiciel Pathway Studio montre une dominance des GO liées au métabolisme lipidique (Annexe 3). Nous avons effectué une dernière recherche d'enrichissement sous PANTHER (Figure 11). Nous constatons également que, dans notre jeu de GDE, nous avons une surreprésentation des GO associées au métabolisme des lipides qui se traduit par des Fold Enrichment élevés. Quant aux autres GO, elles couvrent un champ d'implication très vaste dont celui des processus métaboliques, nous n'en tiendrons donc pas compte.

Cette analyse GO a permis de mettre en évidence la prépondérance du métabolisme lipidique dans notre étude ainsi que la présence de quelques autres processus tels que le métabolisme protéique ou les processus de communication et de transduction de signaux. Au vu de ces résultats et de l'objectif du projet nous nous sommes focalisés sur le métabolisme lipidique lors des analyses suivantes. Il convient maintenant d'analyser plus en détail le rôle des GDE intervenant dans ce métabolisme et les effets que peut avoir leur régulation.



**Figure 10 :** Enrichissement en ontologie des gènes (GO) issu du logiciel ProteINSIDE via le logiciel Cytospace. Les GO enrichies sont encadrées en rouge (Source personnelle d'après ProteINSIDE).

Métabolisme lipidique  
  Régulation des fonctions moléculaires

nb de gènes de cette GO dans notre échantillon  
 ↓  
 valeur attendue si les gènes avaient été choisis aléatoirement

nb de gènes par GO chez l'humain  
 ↓

GO biological process complete	Homo sapiens (REF)	Client Text Box Input (▼ Hierarchy NEW! ?)	#	expected	Fold Enrichment	+/-	P value
<b>fatty acid beta-oxidation</b>	58	8	.66	12.11	+	3.73E-03	
↳ fatty acid catabolic process	75	9	.85	10.54	+	2.25E-03	
↳ fatty acid metabolic process	321	17	3.66	4.65	+	1.93E-03	
↳ monocarboxylic acid metabolic process	589	28	6.71	4.17	+	2.23E-06	
↳ carboxylic acid metabolic process	1098	36	12.50	2.88	+	1.11E-04	
↳ oxoacid metabolic process	1221	36	13.90	2.59	+	1.51E-03	
↳ organic acid metabolic process	1238	36	14.10	2.55	+	2.11E-03	
↳ single-organism cellular process	12003	186	136.67	1.36	+	1.02E-07	
↳ single-organism process	13224	196	150.58	1.30	+	6.01E-07	
↳ cellular process	14730	198	167.72	1.18	+	3.10E-02	
↳ small molecule metabolic process	2660	61	30.29	2.01	+	4.62E-04	
↳ single-organism metabolic process	5337	92	60.77	1.51	+	4.81E-02	
↳ cellular lipid metabolic process	1015	34	11.56	2.94	+	1.76E-04	
↳ lipid metabolic process	1259	39	14.34	2.72	+	1.17E-04	
↳ monocarboxylic acid catabolic process	103	9	1.17	7.67	+	2.97E-02	
↳ carboxylic acid catabolic process	228	16	2.60	6.16	+	1.02E-04	
↳ organic acid catabolic process	228	16	2.60	6.16	+	1.02E-04	
↳ small molecule catabolic process	358	23	4.08	5.64	+	3.33E-07	
↳ single-organism catabolic process	1058	37	12.05	3.07	+	1.24E-05	
↳ cellular lipid catabolic process	164	14	1.87	7.50	+	7.79E-05	
↳ lipid catabolic process	262	14	2.98	4.69	+	2.06E-02	
↳ fatty acid oxidation	76	9	.87	10.40	+	2.51E-03	
↳ lipid oxidation	77	9	.88	10.27	+	2.80E-03	
regulation of localization	2390	53	27.21	1.95	+	1.24E-02	
<b>regulation of catalytic activity</b>	2350	52	26.76	1.94	+	1.71E-02	
↳ regulation of molecular function	2832	59	32.25	1.83	+	2.02E-02	
regulation of biological quality	3602	70	41.01	1.71	+	2.06E-02	
response to stimulus	8117	129	92.42	1.40	+	8.99E-03	

↑  
 valeur d'enrichissement

**Figure 11 :** Liste de l'ontologie des gènes enrichie avec le logiciel Panther (Source personnelle d'après Panther).

### 4.3. ANALYSE DES VOIES METABOLIQUES (PATHWAYS)

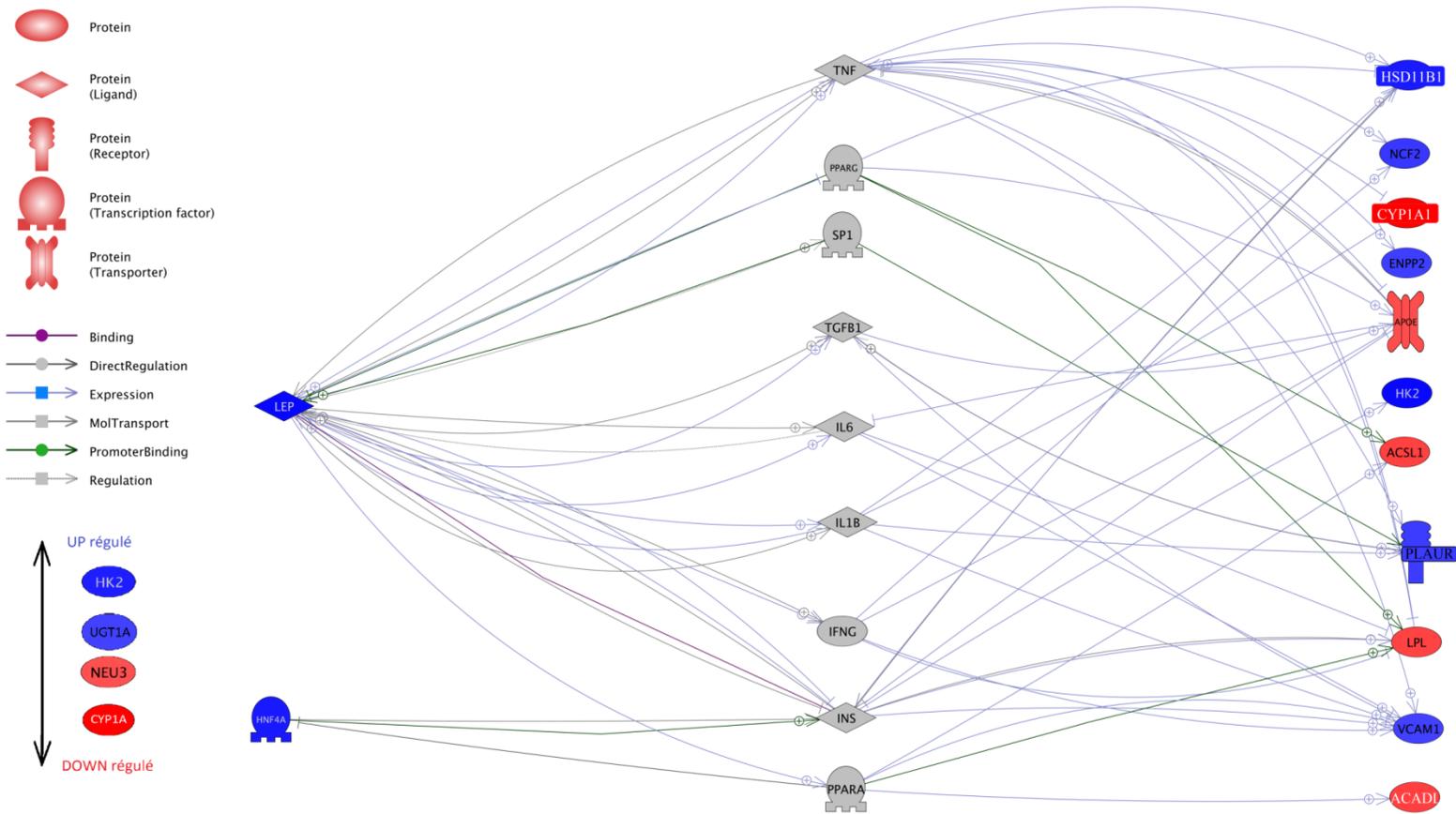
L'analyse des voies métaboliques consiste à identifier les voies cellulaires dans lesquelles interviennent nos GDE. Pour cela nous avons utilisé deux outils : KEGG et Pathway Studio.

KEGG nous a permis d'identifier les voies cellulaires enrichies, parmi lesquelles apparaissent le métabolisme des AG, leur dégradation ainsi que la voie de signalisation PPAR (facteur de transcription très connu pour réguler de nombreux gènes impliqués dans le métabolisme des lipides). Nous observons notamment une surexpression des gènes codant pour des enzymes impliquées dans la bêta-oxydation des AG dans le TA des chèvres lactantes (Annexe 4).

Avec Pathway Studio, nous observons de nombreuses voies lipidiques enrichies telles que la régulation de la lipogenèse dans les adipocytes qui met en évidence notamment une surexpression de gènes codant pour la LPL, l'ACSL1 et la DGAT1 (Annexe 5).

### 4.4. ANALYSE ET CONSTRUCTION DE RESEAUX

En parallèle, j'ai également cherché à identifier des réseaux de régulation, c'est-à-dire établir des interactions entre nos différents gènes. Cette analyse avait pour objectif d'identifier des régulateurs clés du métabolisme lipidique dans notre modèle d'étude. A l'aide de Pathway Studio nous avons obtenu, après filtration des données (gènes sans relation et nombre de références bibliographiques insuffisantes), un réseau contenant 13 GDE et 9 régulateurs non différentiellement exprimés (Figure 12). Ce réseau met en évidence notamment que la leptine (LEP, hormone sécrétée par le TA corrélée positivement à la masse adipeuse, annexe 6) et le facteur de transcription HNF4A (annexe 7), dont les expressions sont diminuées durant la lactation, semblent avoir un rôle important du fait qu'ils régulent des facteurs de transcription tels que les PPAR ainsi que des molécules de signalisation cellulaires comme TGF $\beta$  et IL6.



**Figure 12 :** Réseau des régulateurs communs aux gènes différentiellement exprimés impliqués dans le métabolisme lipidique selon le logiciel Pathway Studio (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle d'après Pathway Studio).

## 4.5. DISCUSSION

La lactation (comparaison des TA des chèvres lactantes contre non lactantes) a entraîné une modification de l'expression de 252 gènes dont 118 surexprimés et 134 sous exprimés. La classification GO de ces 252 GDE a permis de mettre en lumière 25% de gènes (64 GDE) impliqués dans le métabolisme des lipides : ce fort pourcentage est en accord avec les analyses d'enrichissements réalisées. En effet, quel que soit le logiciel utilisé, ce processus biologique ressort systématiquement et en priorité dans nos analyses.

Sur le plan biologique, d'une façon plutôt surprenante, la lactation, dans le TA caprin, semble globalement augmenter l'expression des gènes impliqués dans la synthèse d'AG (ACOT2, PC) et de TG (DGAT1 et DGAT3), la captation (LPL) et l'activation (ACSL1) des AG et diminuer ceux impliqués dans le transport (VLDLR, GM2A, APOE, CRABP2) et le stockage (PLIN4) des lipides. Ces résultats sont inattendus car, *à priori*, en désaccord avec les observations effectuées sur vaches laitières et brebis où la lipogénèse est diminuée et la lipomobilisation augmentée dans le TA d'animaux lactants pour favoriser la production de lait par la glande mammaire<sup>23</sup>. Nous observons également que les gènes (ACAT1, ACADL, ALAD, etc...) impliqués dans la bêta-oxydation des AG (source majeur d'énergie sous forme d'ATP) sont augmentés dans le TA des chèvres laitières. Ces résultats semblent plutôt cohérents avec le fait que ces animaux doivent trouver de l'énergie pour produire du lait (Tableau III).

Cependant avant de tirer des conclusions sur ce travail, il faudra dans un premier temps valider *a minima* 10% des GDE par une approche plus conventionnelle de RT-qPCR, puis, dans un second temps, poursuivre l'exploitation des résultats en les confrontant, notamment, aux données zootechniques. En effet, les données de note d'état corporel, de niveau de production laitière, de niveau d'ingestion, de stade de lactation, ... des animaux doivent être attentivement étudiées pour interpréter le plus justement possible ces résultats et les confronter aux données existantes de la bibliographie chez le ruminant.

---

<sup>23</sup> Chilliard, « Métabolisme du tissu adipeux, lipogénèse mammaire et activités lipoprotéine-lipasiques chez la chèvre au cours du cycle gestation-lactation ».

Néanmoins il faut noter que les annotations ont été réalisées en utilisant le génome bovin (*Bos taurus*). Une nouvelle annotation en utilisant le génome caprin (*Capra hircus*) bien que moins bien renseigné devrait confirmer ou non nos premiers résultats.

**Tableau III :** Liste des gènes impliqués dans le métabolisme lipidique classés par catégorie fonctionnelle (Source personnelle).

Catégories Fonctionnelles	Symbole	protéine	log_ratio	Moy reads_non-lactante	Moy reads_lactante
Synthèse d'acides gras	ACOT2	Acyl-CoA Thioesterase 2	1,63	182	565
	HK2	Hexokinase 2	-2,35	1025	201
	PC	Pyruvate Carboxylase	0,97	3068	6023
Synthèse phospholipide	CDS2	Phosphatidate Cytidylyltransferase 2	-1,01	7988	3968
Synthèse de triglycérides	DGAT1	Diacylglycerol O-Acyltransferase 1	1,25	1463	3484
	GPAT3	Glycerol-3-Phosphate Acyltransferase 3	2,07	1899	7951
Stockage des lipides	PLIN4	Perilipin 4	-1,10	73035	33988
Transport des lipides	VLDLR	Very Low Density Lipoprotein Receptor	-1,91	168	46
	CRABP2	Cellular Retinoic Acid Binding Protein 2	-2,58	2635	440
	ETFDH	Electron-Transferring-Flavoprotein Dehydrogenase	1,08	2375	5004
	GM2A	GM2 Ganglioside Activator	-0,86	1866	1030
	APOE	Apolipoprotein E	1,09	2301	4907
Activation des acides gras	ACSS1	Acyl-CoA Synthetase Short-Chain Family Member	1,10	3478	7437
Captation et transport des acides gras	LPL	Lipoprotein Lipase	1,31	61735	153062
Facteur de transcription	LEP	Leptin	-2,39	48031	9194
	HNF4A	Hepatocyte Nuclear Factor 4, Alpha	-1,85	509	142
	PRKAR2B	Protein Kinase, CAMP-Dependent, Regulatory, Type 2	1,08	2837	5988
Oxydation des acides gras	ACAT1	Acetyl-CoA Acetyltransferase 1	1,02	16960	34359
	ACSL1	Acyl-CoA Synthetase Long-Chain Family Member 1	1,36	40194	103471
	ALAD	Aminolevulinatase Dehydratase	1,20	3209	7395
	CYP1A1	Cytochrome P450, Family 1, Subfamily A, Polypeptide	3,58	13	149
	DECOR1	2,4-Dienoyl CoA Reductase 1, Mitochondrial	1,18	4244	9591
	EHHADH	Enoyl-CoA, Hydratase/3-Hydroxyacyl CoA Dehydrogenase	1,27	1188	2863
	HSD17B10	Hydroxysteroid (17-Beta) Dehydrogenase 10	1,26	3863	9269
	ECI1	Enoyl-CoA Delta Isomerase 1	1,17	2246	5038
	ETFA	Electron-Transfer-Flavoprotein, Alpha Polypeptide	1,12	2213	4815
	SLC25A20	Solute Carrier Family 25 (Carnitine/Acylcarnitine Transporter)	0,93	4094	7801
Régulation métabolisme lipidique via PPARalpha	G0S2	G0/G1 Switch 2	-0,91	8016	4280
Adipogenèse	ADIRF	Adipogenesis Regulatory Factor	-0,93	46574	24399

## CONCLUSION

L'exploitation des données issues du séquençage des ARN haut débit des chèvres lactantes et non lactantes a commencé par une identification des GDE. Ceux-ci ont ensuite été soumis à des analyses d'enrichissement GO afin de les classer et de mettre en évidence les processus biologiques les plus représentés. Le métabolisme lipidique étant largement surreprésenté et sachant que notre sujet d'étude était le TA, nous avons focalisé nos analyses sur ce métabolisme. Nous avons ensuite essayé de trouver des voies cellulaires enrichies avec nos GDE. Nous avons observé que la lipogenèse et l'oxydation des acides gras semblaient être régulées positivement chez la chèvre lactante. Nous avons, en parallèle, cherché à identifier des régulateurs clés en construisant des réseaux de gènes. Ces réseaux ont permis de mettre en avant le rôle joué par la leptine sur des facteurs de transcription tels que les PPAR mais ne permettent pas d'avancer, pour l'instant, de réponse à la surexpression des gènes impliqués dans la lipogenèse. Pour cela il faudrait poursuivre le travail amorcé en exploitant les données de séquençage avec les annotations *Capra hircus* (en cours d'annotation à la plateforme) et mettre en regard les données zootechniques de ces animaux afin d'interpréter au mieux nos résultats.

Le travail effectué a néanmoins permis d'avancer une stratégie d'exploitation de données RNA-seq en utilisant des outils de recherche d'enrichissement de GO et de pathway ainsi que de création de réseaux. Ces outils ont été choisis pour leur facilité de prise en main ainsi que pour l'organisation et la cohérence de leurs résultats, mais aussi pour leur disponibilité (gratuité ou licences payées par la Plate-forme d'Exploration du Métabolisme de l'Inra Theix). Il serait souhaitable de continuer cette recherche afin de découvrir des logiciels potentiellement plus efficaces que ceux évoqués, tels que Metacore, Ingenuity et Netwalker, et de rester informé d'éventuelles mises à jour des outils existants afin de s'adapter et de garder une stratégie d'exploitation efficace.

# BIBLIOGRAPHIE

- Agreste. « Enquête annuelle laitière 2014 ». Gouvernementale. *Ministère de l'agriculture, de l'agroalimentaire et de la forêt - La statistique, l'évaluation et la prospective agricole*, 2015.
- Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. « Gene Ontology: Tool for the Unification of Biology ». *Nature Genetics* 25, n° 1 (mai 2000): 25-29.
- Auvergne Sciences. « Annuaire - INRA - Institut National de la Recherche Agronomique ». *Auvergne Sciences*, 25 février 2011.
- Bonnet, Agnès, Sandrine Lagarrigue, Laurence Liaubet, Christèle Robert-Granié, Magali SanCristobal, et Gwenola Tosser-Klopp. « Pathway results from the chicken data set using GOTM, Pathway Studio and Ingenuity softwares ». *BMC Proceedings* 3, n° 4 (2009): 1-6.
- Bonnet, Muriel, et Yves Chilliard. « Régulation par l'état nutritionnel et la photopériode de l'expression de gènes spécifiant des enzymes du métabolisme énergétique dans le tissu adipeux et le muscle de brebis. » Text, Limoges, 1999.
- Chilliard, Yves. « Métabolisme du tissu adipeux, lipogénèse mammaire et activités lipoprotéine-lipasiques chez la chèvre au cours du cycle gestation-lactation ». Text, Paris 6, 1985.
- Faulconnier, Y., Y. Chilliard, M. B. Montazer Torbati, et C. Leroux. « The Transcriptomic Profiles of Adipose Tissues Are Modified by Feed Deprivation in Lactating Goats ». *Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics* 6, n° 2 (juin 2011): 139-49.
- Faulconnier, Yannick, Carole Delavaud, et Yves Chilliard. « Effets de l'insuline et/ou de la dexaméthasone sur les activités enzymatiques et métaboliques des tissus adipeux bovin et ovin maintenus en culture durant 7 jours ». Limoges, 1998.
- INRA. « Jalons historiques », 16 janvier 2013. « Les chiffres clés », 18 janvier 2013. <http://institut.inra.fr/Reperes/Chiffres>. « Présentation du Centre », 18 janvier 2013.
- J. Delafontaine. « L'analyse de données RNA-seq: mode d'emploi », bioinfo-fr.net.

- Jenssen, Tor-Kristian, Astrid Lægreid, Jan Komorowski, et Eivind Hovig. « A Literature Network of Human Genes for High-Throughput Analysis of Gene Expression ». *Nature Genetics* 28, n° 1 (mai 2001): 21-28.
- Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, et Mao Tanabe. « KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets ». *Nucleic Acids Research*, 10 novembre 2011.
- Kaspric, Nicolas, Brigitte Picard, Matthieu Reichstadt, Jérémy Tournayre, et Muriel Bonnet. « ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses ». *PLOS ONE* 10, n° 5 (22 mai 2015).
- Khan, Sami M., Ole-Petter R. Hamnvik, Mary Brinkoetter, et Christos S. Mantzoros. « Leptin as a Modulator of Neuroendocrine Function in Humans ». *Yonsei Medical Journal* 53, n° 4 (1 juillet 2012): 671-79.
- Le Crom, S., J.-R. Claude, Marzin, et Abdon. « Concept paper : Le séquençage a haut débit méthodes et enjeux en médecine, pharmacologie et toxicologie ». afssaps, 2011.
- Leroux, Christine, Christian Beauvallet, Laurence Bernard, Claudia Bevilacqua, Frédéric Bouvier, Cédric Cabau, Christelle Cebo, et al. « MilkChEST, an integrative approach of the secretion pathways in the mammary epithelial cell: molecular dissection of milk fat adaptability ». In *European Regional Conference On Goats*, np. Tromsø, Norway, 2013.
- Nikitin, Alexander, Sergei Egorov, Nikolai Daraselia, et Ilya Mazo. « Pathway Studio—the Analysis and Navigation of Molecular Networks ». *Bioinformatics* 19, n° 16 (1 novembre 2003): 2155-57.
- Oshlack, Alicia, Mark D. Robinson, et Matthew D. Young. « From RNA-Seq Reads to Differential Expression Results ». *Genome Biology* 11, n° 12 (2010): 220.  
« Pathay Studio - Fact Sheet ». Elsevier Inc., 2015.
- Schibler, L., D. Vaiman, A. Oustry, C. Giraud-Delville, et E. P. Cribiu. « Comparative Gene Mapping: A Fine-Scale Survey of Chromosome Rearrangements between Ruminants and Humans ». *Genome Research* 8, n° 9 (septembre 1998): 901-15.
- Sumner-Thomson, J. M., J. L. Vierck, et J. P. McNamara. « Differential expression of genes in adipose tissue of first-lactation dairy cattle ». *Journal of Dairy Science* 94, n° 1 (janvier 2011): 361-69.

Vernon, Richard G. « Lipid metabolism during lactation: a review of adipose tissue-liver interactions and the development of fatty liver ». *Journal of Dairy Research* 72, n° 4 (novembre 2005): 460–469.

## TABLE DES SIGLES ET/OU DES ABREVIATIONS

ACADL: *Acyl-CoA dehydrogenase, long chain*

ACAT1 : *Acetyl-CoA Acetyltransferase 1*

ACOT2: *Acyl-CoA Thioesterase 2*

ACSL1 : *Acyl-CoA Synthetase Long-Chain Family Member 1*

AG : *Acide Gras*

ALAD: *Aminolevulinate Dehydratase*

ANR : *Agence Nationale de la Recherche*

APOE : *Apolipoprotein E*

ARA: *Auvergne Rhones-Alpes*

BIOM : *Equipe Biomarqueurs*

BP : *Biological process*

CC : *Cellular component*

CRABP2 : *Cellular Retinoic Acid Binding Protein 2*

DGAT1 : *Diacylglycerol O-Acyltransferase 1*

DGAT3: *Diacylglycerol O-Acyltransferase 3*

FC : *Foldchange*

GDE : *Gènes Différentiellement Exprimés*

GM2A: *GM2 Ganglioside Activator*

GO : *Gene Ontology*

HNF4A: *Hepatocyte Nuclear Factor 4, Alpha*

IL6 : *Interleukin 6*

Inra : *Institut National de la Recherche Agronomique*

INS : *Insulin*

LEP : *Leptin*

LHS : *Lipase hormono-sensible*

LPL : *Lipoprotein Lipase*

MF : *Molecular function*

NGS : *Next Generation Sequencing*

PC: *Pyruvate Carboxylase*

PCR : *Polymerase chain reaction*

PLIN4 : *Perilipin 4*

PPAR : *Peroxisome proliferator activated receptor*

RIN : *Rna Integrity Number*

RNA-seq : *RNA-sequencing*

RT-qPCR : *Reverse Transcription- Real-Time Polymerase Chain reaction*

TA : *Tissu Adipeux*

TG : *Triglycéride*

TGFb : *Tumor Necrosis Factor Beta*

UMRH : *Unité Mixte de Recherche sur les Herbivores*

VLDLR : *Very Low Density Lipoprotein Receptor*

## TABLE DES TABLEAUX ET DES FIGURES

<b>Figure 1</b>	Carte des implantations du Centre Inra Auvergne-Rhône-Alpes (Plaquette des activités du centre de recherche Auvergne-Rhône-Alpes, mai 2015).
<b>Figure 2</b>	Représentation schématique et photo au microscope de cellules adipeuses (Anatomy & Physiology par OpenStax, chapitre 4.3, 2013).
<b>Figure 3</b>	Molécule de Triglycéride (Roger DANIELS, 2010).
<b>Figure 4</b>	Schéma récapitulatif du métabolisme des lipides dans le tissu adipeux de ruminant (Y. Chilliard, 1985).
<b>Figure 5</b>	Phase d'amplification par PCR par pont durant un séquençage Illumina. (The Illumina HiSeq 2000 Sequencing Technology, German Cancer Research Center, 2015).
<b>Figure 6</b>	Phase de séquençage par excitation laser des fluorochromes durant un séquençage Illumina. (The Illumina HiSeq 2000 Sequencing Technology, German Cancer Research Center, 2015).
<b>Figure 7</b>	Répartition des gènes différentiellement exprimés selon leurs ontologies obtenues avec le logiciel PANTHER (Source personnelle d'après Panther et VisiBone).
<b>Figure 8</b>	Répartition des gènes différentiellement exprimés selon leurs ontologies après analyse personnelle (Source personnelle).
<b>Figure 9</b>	Pourcentage de gènes différentiellement exprimés dans les processus métabolismes lipidique, glucidique et protéique (Source personnelle).
<b>Figure 10</b>	Enrichissement en ontologie des gènes (GO) issu du logiciel ProteINSIDE via le logiciel Cytospace. Les GO enrichies sont encadrées en rouge (Source personnelle d'après ProteINSIDE).
<b>Figure 11</b>	Liste de l'ontologie des gènes enrichie avec le logiciel Panther (Source personnelle d'après Panther).
<b>Figure 12</b>	Réseau des régulateurs communs aux gènes différentiellement exprimés impliqués dans le métabolisme lipidique selon le logiciel Pathway Studio (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle d'après Pathway Studio).
<b>Tableau I</b>	Concentration et qualité des ARNm extraits des tissus adipeux omental des chèvres lactantes et non lactantes. (Source personnelle).
<b>Tableau II</b>	Gènes les plus exprimés parmi les gènes différentiellement exprimés (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle).
<b>Tableau III</b>	Liste des gènes impliqués dans le métabolisme lipidique classés par catégorie fonctionnelle (Source personnelle).

# TABLES DES ANNEXES ET ANNEXES

**Annexe 1 :** Pourcentage de gènes sur-exprimés (*UP*) et sous-exprimés (*DOWN*) par ontologie (Source personnelle).

**Annexe 2 :** Pourcentage des gènes sur-exprimés (*UP*) et sous-exprimés(*DOWN*) dans les processus métabolismes lipidique, glucidique et protéique (Source personnelle).

**Annexe 3:** Liste de gènes ontologies (*GO*) enrichies selon le logiciel Pathway Studio (Source personnelle).

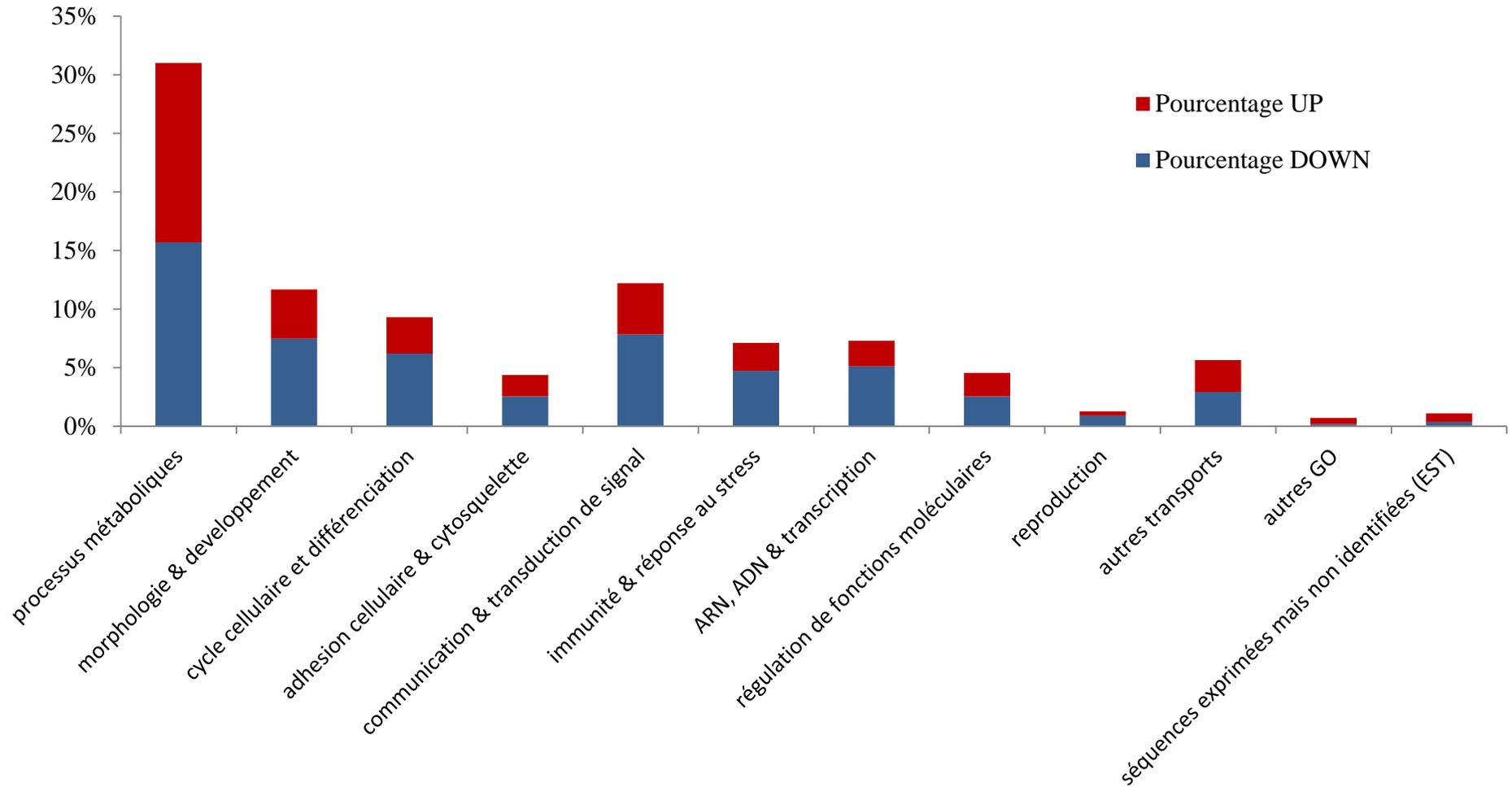
**Annexe 4 :** Voie de la dégradation des acides gras selon le logiciel KEGG (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle d'après KEGG).

**Annexe 5 :** Voie cellulaire de la régulation de la lipogenèse dans les adipocytes d'après Pathway Studio (Source personnelle d'après Pathway Studio).

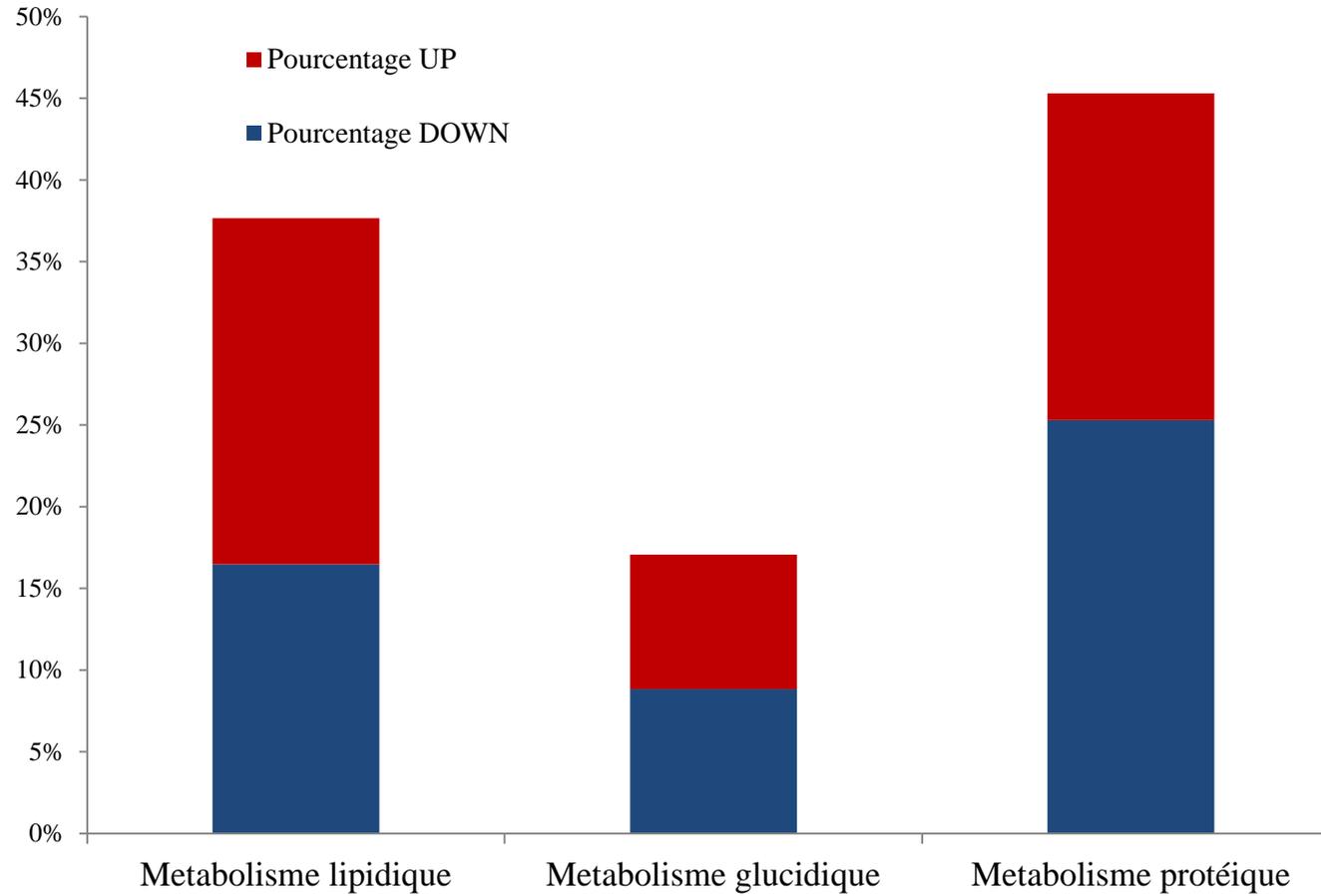
**Annexe 6 :** Réseau des gènes différentiellement exprimés ciblés indirectement par la leptine via PPARs et IL6 d'après Pathway Studio (Source personnelle d'après Pathway Studio).

**Annexe 7 :** Réseau des gènes différentiellement exprimés ciblés indirectement par le facteur de transcription HNF4A via PPARA, INS et IL6 d'après Pathway Studio (Source personnelle d'après Pathway Studio).

**Annexe 1 :** Pourcentage de gènes sur-exprimés (UP) et sous-exprimés (DOWN) par ontologie (Source personnelle).



**Annexe 2** : Pourcentage des gènes sur-exprimés (UP) et sous-exprimés(DOWN) dans les processus métabolismes lipidique, glucidique et protéique (Source personnelle).

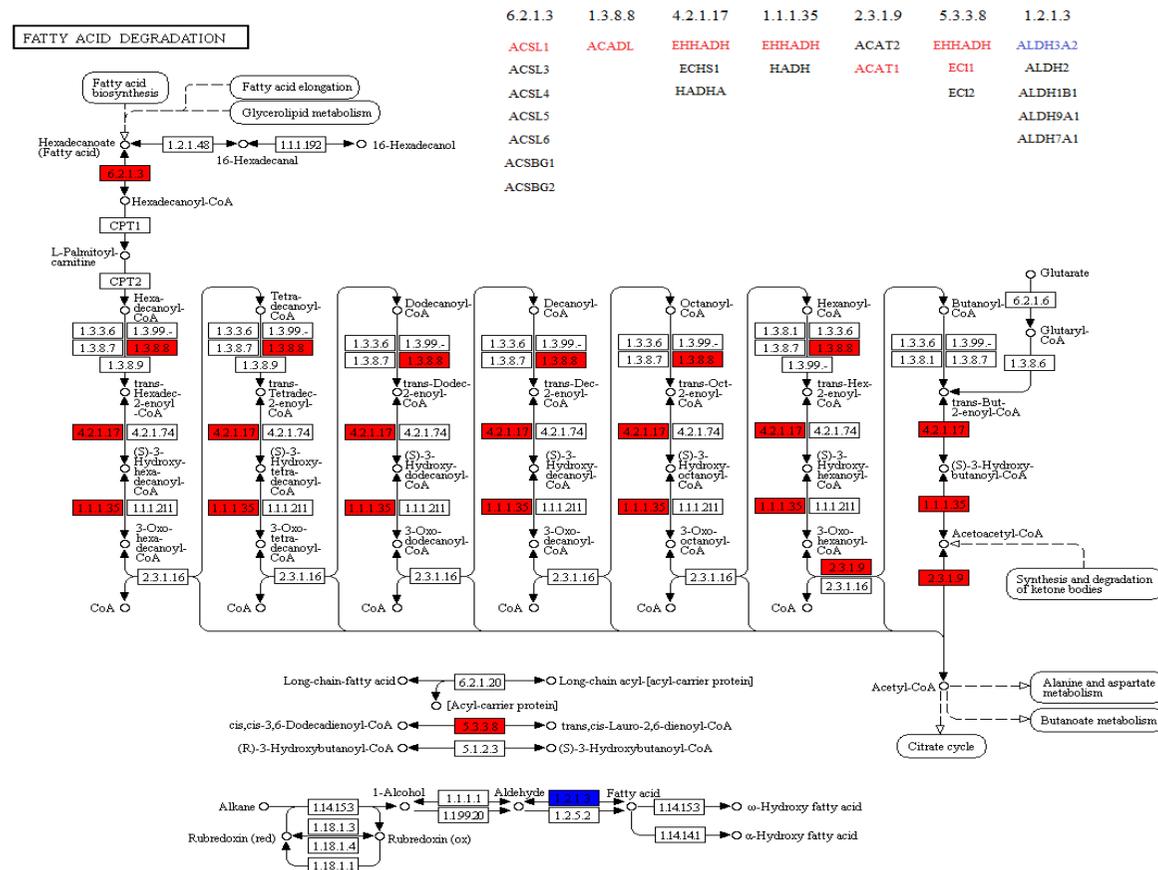


**Annexe 3:** Liste de gènes ontologies (GO) enrichies selon le logiciel Pathway Studio (Source personnelle).

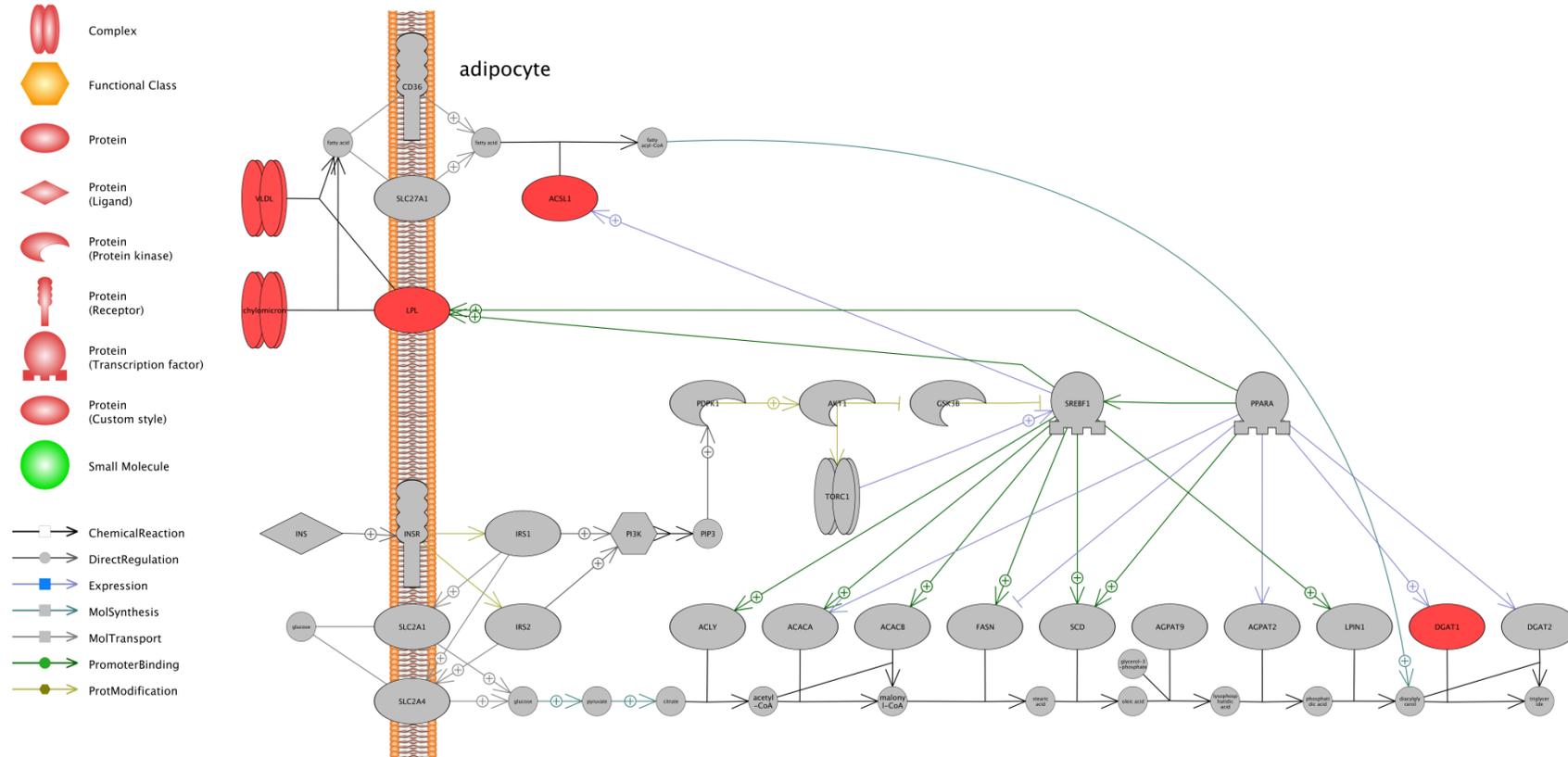
Name	# of Entities	Overlap	Percent Ove	Overlapping Entitie	p-value
response to hormone	101	8	7	HCN2;TIMP2;VLDLR	5,8001E-06
lipid metabolic process	506	17	3	AGPAT9;HSD11B1;C	6,2082E-06
response to drug	509	17	3	UGT1A1;HCN2;ACSL	6,7087E-06
small molecule metabolic process	1414	31	2	ACSL1;APOE;SLC22	8,5687E-06
cellular lipid metabolic process	156	9	5	AGPAT9;DGAT1;AC	1,7126E-05
branched-chain amino acid catabolic process	18	4	22	DBT;HSD17B10;MC	2,0782E-05
fatty acid beta-oxidation using acyl-CoA dehydrogenase	7	3	42	ACADL;ETFDH;IVD	2,7726E-05
acyl-CoA metabolic process	28	4	14	EHHADH;ACOT2;GL	0,00012921
response to organic cyclic compound	253	10	3	EDNRB;UGT1A1;AC	0,00014552
aging	254	10	3	EDNRB;HCN2;APOE	0,00015027
mesenchymal cell differentiation	12	3	25	WWTR1;PLAUR;NO	0,00016834
very-low-density lipoprotein particle clearance	4	2	50	APOE;VLDLR	0,00051947
metabolic process	1301	25	1	UGT1A1;PFKFB4;HK	0,00052376
response to organic substance	153	7	4	UGT1A1;DMBT1;AC	0,0006167
triglyceride biosynthetic process	42	4	9	AGPAT9;DGAT1;AC	0,00063751
response to lipopolysaccharide	252	9	3	EDNRB;UGT1A1;CY	0,0006505
porphyrin-containing compound metabolic process	20	3	15	UGT1A1;CYP1A1;AL	0,00082532
response to hypoxia	259	9	3	CYP1A1;VLDLR;APC	0,0008344
insemination	5	2	40	TAC1;DDO	0,00086042
AMPA glutamate receptor clustering	5	2	40	APOE;DLG4	0,00086042
leucine catabolic process	5	2	40	MCCC1;IVD	0,00086042
fatty acid beta-oxidation	46	4	8	ACADL;ECI1;DECR1	0,00090269
triglyceride metabolic process	46	4	8	DGAT1;ACSL1;APO	0,00090269
semaphorin-plexin signaling pathway	21	3	14	PLXNA4;PLXNA1;SE	0,00095624
regulation of smooth muscle cell migration	6	2	33	ACE;PLXNA1	0,00128265
cell-substrate junction assembly	6	2	33	TNS1;TLN1	0,00128265
negative regulation of membrane protein ectodomain proteolysis	6	2	33	TIMP2;TIMP3	0,00128265
negative regulation of metalloenzyme activity	6	2	33	TIMP2;TIMP3	0,00128265
response to nutrient	139	6	4	UGT1A1;ACSL1;VCA	0,00202648
negative regulation of mitotic cell cycle	28	3	10	NKX3-1;TIMP2;HNF	0,00224441

metabolisme	nb de Go	overlap
lipid	9	56
protein	3	8
cell cycle	3	8
communication	2	5
reproduction	1	2
regulation of molecular function	1	2

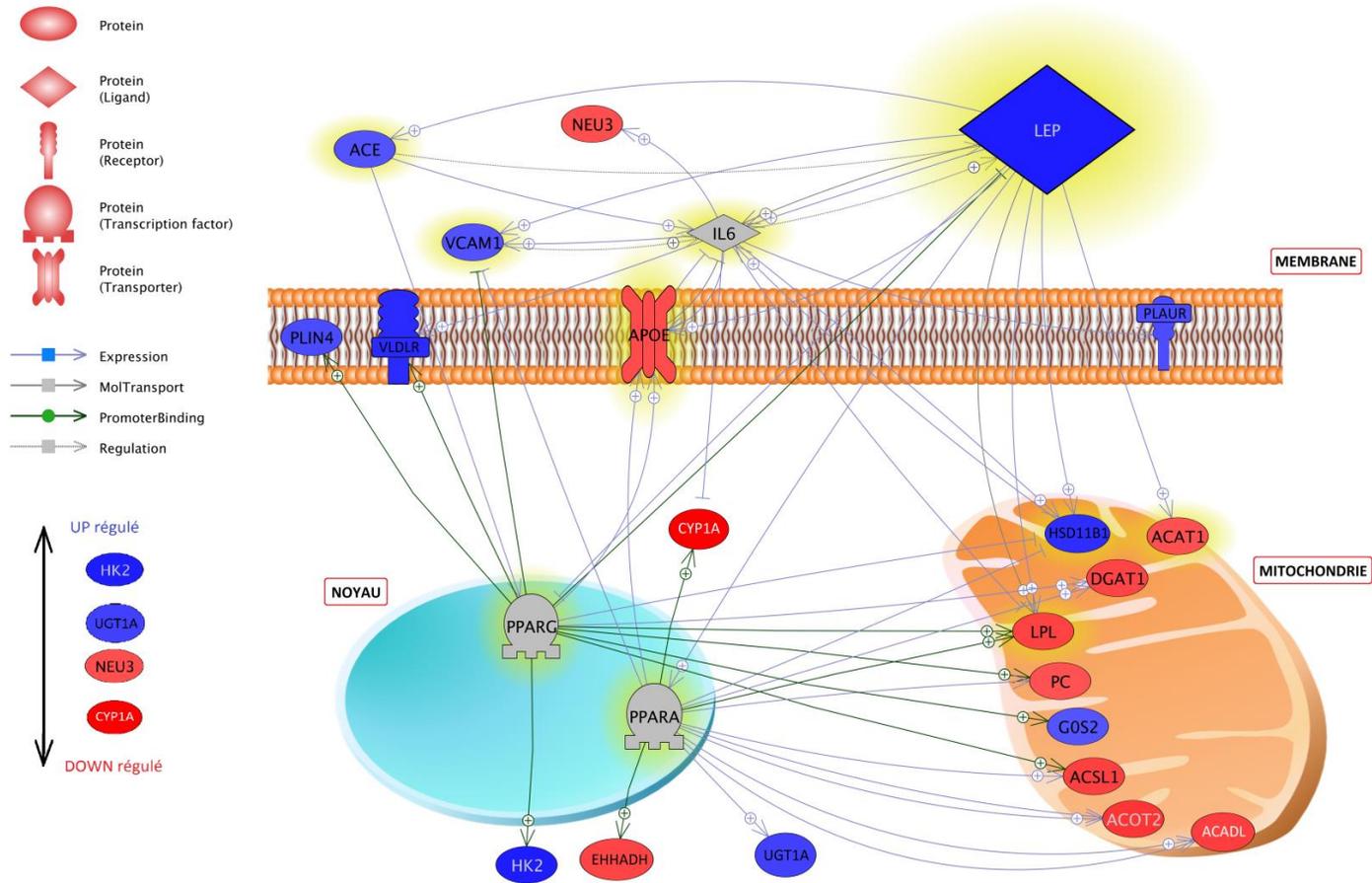
**Annexe 4 :** Voie de la dégradation des acides gras selon le logiciel KEGG (en rouge les gènes sur-exprimés et en bleu ceux sous-exprimés, source personnelle d'après KEGG).



**Annexe 5 :** Voie cellulaire de la régulation de la lipogenèse dans les adipocytes d'après Pathway Studio (Source personnelle d'après pathway Studio).



**Annexe 6 :** Réseau des gènes différentiellement exprimés ciblés indirectement par la leptine via PPARs et IL6 d'après Pathway Studio (Source personnelle d'après Pathway Studio).



**Annexe 7 :** Réseau des gènes différentiellement exprimés ciblés indirectement par le facteur de transcription HNF4A via PPARA, INS et IL6 d'après Pathway Studio (Source personnelle d'après Pathway Studio).

