N°

/__/__/__/__/__/__/__/__/__/__/

THÈSE

pour obtenir le grade de

Docteur

de

l'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech)

Spécialité : Génétique

présentée et soutenue publiquement par

Félicie FAUCON

le 26 mars 2009

IDENTIFICATION, CHEZ LES RUMINANTS, DES GENES OU RESEAUX DE GENES IMPLIQUES DANS LA DIFFERENCIATION ET LE FONCTIONNEMENT DE LA GLANDE MAMMAIRE

sous la direction de : ***Patrice MARTIN***

à l'INRA de Jouy-en-Josas, dans l'équipe Génomique Expressionnelle & Lait (GEL-PiCT) au sein de l'Unité Génomique et Physiologie de la Lactation (UR1196)

Devant le jury :

M. Jean DJIANE, DR, **INRA** …….. ……..	**Rapporteur**
M. Armand SANCHEZ, Pr, **FACULTAT DE VETERINARIA** ……..	**Rapporteur**
M. Thomas HEAMS, MC, **AgroParisTech** …….. ……..	**Président**
Mme Christine LEROUX, CR, **INRA** …….. ……..	**Examinateur**
Mme Sophie MATTALIA, Ingénieur, **INSTITUT DE L'ELEVAGE**……	**Examinateur**
M. Patrice MARTIN, DR, **INRA** …….. ……..	**Examinateur**

L'Institut des Sciences et Industries du Vivant et de l'Environnement (Agro Paris Tech) est un Grand Etablissement dépendant du Ministère de l'Agriculture et de la Pêche, composé de l'INA PG, de l'ENGREF et de l'ENSIA (décret n° 2006-1592 du 13 décembre 2006)

Remerciements

Aux membres du jury de thèse :

Monsieur Jean Djiane, directeur de recherche à l'INRA de Jouy-en-Josas. Merci d'avoir accepté de lire et corriger cette thèse. Merci pour vos nombreuses remarques qui ont ouvert la réflexion sur de nombreuses parties de ce travail.

El señor Armand Sanchez, un profesor de la facultat de Veterinaria en Barcelona. Muchas gracias por aceptar leer la tesis y darla una amplitud international.

Monsieur Patrice Martin, directeur de thèse, responsable de l'équipe Génomique Expressionnel et Lait (GEL, nouvellement Lait Génomique Santé, LGS) de l'unité Génomique et Physiologie de la Lactation (GPL) de l'INRA de Jouy-en-Josas. Merci pour l'accueil au sein de ton laboratoire et l'appui dans la réalisation de ce travail.

Monsieur Thomas Heams, président du jury, tuteur d'AgroParisTech et membre du comité de thèse. Merci pour ta disponibilité, les précieux conseils apportés et surtout merci d'avoir présidé le jury de cette thèse.

Madame Christine Leroux, également membre du comité de thèse. Merci pour le suivi et l'aide apportée aux travaux réalisés, notamment au moment des prélèvements effectués sur les vaches.

Madame Sophie Mattalia, responsable du service Aptitude et Sélection des Race Laitières de l'Institut de l'Elevage, auquel j'étais rattachée par l'intermédiaire de la bourse Cifre ; également membre du comité de thèse. Merci pour la confiance accordée dès le début de ces travaux, le suivi et le soutien actif.

Aux membres du comité de thèse (non cités ci-dessus) :

Jean-Michel Chardigny, Jean-Louis Dacheux, Koenraad Duhem, Pierre-Louis Gastinel, Pascale Leroy et Jean-Claude Mocquot.
Merci pour votre présence, vos nombreux conseils qui m'ont toujours permis de rebondir dans l'avancée de ces travaux. Les discussions lors des différentes réunions ont été très riches pour moi grâce à l'ouverture d'esprit qu'elles apportaient.

Aux collègues de l'équipe INRA dans laquelle j'ai réalisé cette thèse :

Et plus particulièrement,

Emmanuelle Rebours, pour tout ce qu'elle a donné, tant sur le plan professionnel (nocturnes intempestives, mise en relation avec diverses personnes pour la réalisation du travail, mise au point des protocoles, lecture des écrits…) que sur le plan amical (les heures au labo, ça crée des liens !). Travailler à ses côtés a été un réel plaisir, de par son dynamisme, son sourire, ses blagounettes, son écoute et toujours sa bonne humeur. Merci Emma pour ce que tu es, pour ces trois années écoulées et merci pour cette amitié grandissante.

Sophie Pollet, infiniment patiente, elle a été en mesure de m'expliquer dix fois les techniques de blast et de design des amorces, sans se lasser. Toujours présente en cas de coup dur, merci pour les petites pierres apportées à ce travail.

Christian Beauvallet, toujours disponible, même en vacances, il n'a jamais été avare du moindre conseil ! Un seul regret dans cette thèse, ne pas avoir fait de protéomique avec lui. Merci d'avoir toujours su me consacrer du temps, notamment dans la relecture et la correction du manuscrit et d'avoir enrichi mes quelques connaissances en protéomique.

Suzana Krupova, soleil du laboratoire, je ne l'ai jamais vu sans sourire et pourtant j'en ai partagé des jours à ses côtés. Merci d'avoir toujours été la petite oreille pour écouter et le rayon de lumière d'une journée.

Samira Makzhami, discrète mais il faut que les choses se fassent ! Merci d'avoir su me conseiller à chaque exposé en anglais et merci pour les analyses PCR effectuées.

Christelle Cebo, merci d'avoir toujours pris du temps pour me donner tes petits tuyaux et merci pour tes petites blagues avant l'oral d'un congrès !

Claudia Bevilacqua, merci pour toutes les analyses PCR qui représente une masse de travail considérable.

Jean-Christophe Helbling, merci pour tout le travail réalisé dans la partie chèvres gestantes, en y mettant le nez je me suis rendu compte de l'ampleur de ce que tu avais fait.

Guy Miranda, merci pour l'analyse des laits du programme Genomilk Fat.

Sylvain Bellier, colloc de bureau, merci de m'avoir tout simplement supportée ces 3 ans… J'ai remarqué toutefois que tes visites se raréfiaient lors de la rédaction… !

Nisrine Bouhmarou, cuisinière hors pair ! C'est autour d'un bon tajine préparé par Nisrine que l'on ressent toute la chaleur marocaine.

Mais aussi merci à Sanda Andrei, Leonardo Biancchi et Marie-Madeleine Delage avec qui j'ai passé quelques temps et qui ont contribué de plus ou moins loin au travail de cette thèse.

Aux collègues de l'unité GPL :

Michèle Ollivier-Bousquet et Eve Devinoy qui se sont succédées à la direction de l'unité. Merci pour votre accueil au sein de votre unité.

Marie-Elisabeth Marmillod et Martine Palis, assistantes toujours au poste pour répondre aux thésards un peu innocents et toujours en retard !

Laurent Galio pour son coup de pouce dans la mise au point des protocoles, Etienne Aujean, Gisèle Barrière, Eric Chanat, Madia Charlier, Sophie Chat, Stephanie Droineau, Catherine Hue-Beauvais, Clémence Kress, Hichem Lahouassa, Christine Longin, Alain Pauloin, Barbara Petridou, Bagher Torbati et Sandrine Truchet. Merci d'avoir été là, de m'avoir écoutée lors de diverses présentations et d'avoir toujours partagé vos connaissances, votre savoir et savoir-faire.

Aux collègues de l'Institut de l'Elevage :

Jean-Pierre Balansac, Betty Basset, Chantal Bertolotto, Véronique Bonnement, Xavier Bourrigan, Mickaël Brochard, Virginie Clément, Stéphanie Coyral-Castel, Coralie Danchin-Burge, Martine Deredec, Vigan Dervishi, Michel Douget, Delphine Duclos, Marie-Noëlle Fouilloux, Christine Gadenne, Alain Gastebled, François Guillaume, Hélène Leclerc, Martine Marie, Pascale Le Mézec, Lucie Markey, Stéphanie Minéry, Sophie Moureaux, Didier Regaldo, Gilles Thomas, Denis Troboa et tous les autres.
Merci pour vos petites attentions, votre partage, écoute et conseils lors des diverses rencontres que nous avons eu.

A tous les collaborateurs et partenaires des projets :

Julie Aubert et Stéphane Robin de l'UMR 518 AgroParisTech/INRA (Paris), pour l'immense aide apportée dans l'analyse des données de transcriptomique. Vous m'avez permis d'acquérir une certaine autonomie qui a été très bénéfique.

Hélène Larroque de la SGQA (INRA de Jouy-en-Josas), pour tout le soin qu'elle a apportée dans la sélection des animaux, leur transport entre le Pin-au-Haras et Theix et la transmission d'information. Ce que nous avons vécu toutes les deux, en chemin vers le Pin, a représenté un sacré virage dans ma vie.

Laurence Bernard, Yannick Faulconnier et Séverine Ollier et toute l'équipe TALL de l'URH (INRA de Theix), pour leur accueil chaleureux au sein de leur laboratoire lors des biopsies. Merci à Laurence pour l'organisation et la réalisation des biopsies sur les vaches et pour les nombreuses discussions que nous avons eu sur mes données.

Olivia Ménard du STLO (Rennes), pour sa rapidité à répondre et sa co-opération dans les analyses granulométriques sur les laits. Travailler avec toi en DEA a été un réel plaisir, rien n'a changé pour la thèse.

Christèle Robert-Granié de la SAGA (Toulouse), pour sa disponibilité, malgré un emploi du temps très chargé, et ses excellents conseils pour analyser les données du programme Genomilk Fat.

Cedric Cabau, Philippe Bardoux et toute l'équipe de SIGENAE que je n'ai cessé d'enquiquiner pour annoter les lames et publier les données. Merci pour votre efficacité, votre patience et vos réponses si rapides.

Didier Goidin, Caroline Thureau et Louis-Philippe Mars de la société Agilent Technologies pour m'avoir permis de tester leur matériel, cela a grandement enrichi mon travail. Merci pour votre aide et votre sympathie.

Au personnel de l'UCEA, de l'unité de sélection du Pin au Haras et de la ferme expérimentale de Theix pour leur travail remarquable dans le soin apporté au animaux et leur rigueur dans les prélèvements.

A ma famille :

Mes parents, toujours en soutien malgré mes choix de vie. Merci pour votre présence, votre respect, votre écoute et vos conseils. Merci de m'avoir toujours suivie, quelque soit la décision engagée, le choix posé et la difficulté à surmonter. Merci pour votre ouverture d'esprit et vos conseils si précieux.

Hugo, l'aîné de nous tous…l'exemple, le 1er docteur de la fratrie… Merci d'avoir toujours dédramatisé les situations et d'avoir toujours ouvert ta porte à ta petite sœur.

Timothée, un « non médecin », enfin ! Quel bonheur de pouvoir partager avec toi quelques sujets de math ou physique, lorsque les dîners familiaux dérivaient un peu trop sur la médecine ! Merci pour ton accueil dans les premières années de cette thèse et d'avoir fait de mes 20m² un bel espace de vie.

Xavière, et oui, tu le disais si bien, toutes les choses ont une fin…même la thèse ! Merci d'être toujours là quelque soit les circonstances…même si je ne donne pas autant de nouvelles que souhaité. Merci pour ta douceur et ton écoute qui ont toujours été un exemple pour moi.

Tiphaine, ma première « petite sœur ». Après m'avoir vue bien trimer, tu ne prendras peut être pas la direction du doctorat et peut être que cela sera plus sage. Je ne désire que te souhaiter beaucoup de bonheur dans ta vie personnelle et professionnelle.

Damascène, mon petit bonhomme qui a bien grandit ! Merci pour ta compréhension et ton écoute… Beaucoup de caps difficiles à passer durant cette thèse et tu as toujours été là, à l'écoute… Une seule chose : éclate toi dans la vie, trouve ta voie, engouffre toi dedans et profite !

Elise-Péroline, la « petite dernière ». Merci pour ta chaleur lorsque je rentre dans la maison parentale. Tu es celle présente mais aussi celle qui accueille, sans toi rien n'est pareil. Toujours très ambitieuse, je te souhaite de prendre les mesures nécessaires pour réaliser tes rêves les plus secrets. Tiens bon et sois heureuse !

Mon parrain, oncle Michel et ma tante Marie-Madeleine pour leur accueil dans leur foyer alors que je débutais les premiers mois de ce travail. Merci de m'avoir suivi tout au long de ma vie et d'avoir toujours ouvert votre porte et votre cœur dès que j'en avais besoin. Oncle Michel n'aura pas vu l'aboutissement de ces travaux mais il aura été une grande source de réconfort pour les mener à bien. Merci de continuer à veiller sur moi.

Benoît, ami d'hier, fiancé d'aujourd'hui…mari de demain ? Merci de m'avoir attendue et de m'avoir aidée à clôturer ce travail. Merci pour ton soutien, ta bonne humeur et ta chaleur. Un seul désir aujourd'hui, rester à tes côtés aussi longtemps que la vie nous le permettra.

A tous les oncles et tantes, cousins et cousines, petits-cousins et petites-cousines qui m'ont soutenue dans ce projet et dans ma vie quotidienne.

A mes amis :

De la petite enfance, Catherine C., Delphine L., Marie M., Marie C., Agnès C., Jean-Philippe B., Sébastien V. Merci d'avoir toujours respecté mon choix de quitter Lyon et d'avoir

toujours été là lors de mes retours. Amitié si précieuse qu'elle dépasse (bien trop souvent ?) le temps et les frontières.

De l'après-bac, Claire G, Etienne G., Sophie G., Agnès D., Paul R., Anna Z., Laura W. Merci pour les fous rires en toutes occasions, vos visites là où je me trouve et votre accueil lorsque j'ai besoin d'une pause.

De la thèse, Ségolène F., Cécile S., Sophie P., Diane dB, Aliénor dH, Elise & Olivier L., Jerôme dL, Alain R., Marc dL, Marc J., Jerôme dM, Florent S., Gwenaël J., Mathieu J., Valérie B., Mathilde C., père Olivier P., Mr et Mme Casal et tous les autres. Merci d'avoir accepté mes emplois du temps un peu chaotiques et d'être restés tout proche dans les joies comme dans les coups durs !

A mes chers élèves, Ronan, Marie, Florence, Séverine, Camille, Laure, Julie, Ariane, Margaux, Maylis, Gaétan, Guillaume et tous les autres, j'ai semé, à vous de faire fructifier, gardez confiance en vous !

Monsieur Casal pour la réalisation des cinq premières figures de ce manuscrit. Merci d'avoir su reproduire sur simples explications le fond de ma pensée. L'émotion était intense lorsque je découvrais vos schémas et que je remarquais…que vous aviez lu dans mes pensées !

Cette thèse présente une partie du programme Genomilk Fat co-financé par l'ANR (Genanimal) et Apis-Gene. Le projet de thèse a également reçu un support financier du Ministère de l'Agriculture (CAS DAR). Ce travail de thèse a été réalisé en convention CIFRE entre l'Institut de l'Elevage et l'INRA de Jouy-en-Josas.

« Je ne vois qu'un moyen de savoir jusqu'où on peut aller :
c'est de se mettre en route et de marcher » Henri Bergson

Tables des matières

Remerciements ___ ***2***

Tables des matières ___ ***7***

Liste des abréviations ___ ***9***

Liste des Tableaux ___ ***11***

Liste des Figures ___ ***13***

Contexte ___ ***16***

I. Chapitre 1 : Synthèse bibliographique ___ ***20***

I.1. Description morphologique et anatomique de la glande mammaire des ruminants ___ **20**

I.1.1. L'unité de production du lait : la cellule épithéliale mammaire ___ 22

I.1.2. Le tissu sécrétoire ___ 23

I.1.3. Les canaux galactophores ___ 25

I.1.4. La citerne de la glande ___ 26

I.1.5. Les trayons ___ 26

I.1.6. Les mécanismes d'évacuation du lait ___ 28

I.2. Développement et différenciation de la glande mammaire des ruminants ___ **28**

I.2.1. Descriptions morphologiques du développement de la glande mammaire des ruminants ___ 28

I.2.2. Régulation hormonale des différentes étapes de la biologie de la glande mammaire ___ 35

I.3. Composition et biosynthèse des constituants majeurs du lait ___ **42**

I.3.1. Les glucides ___ 43

I.3.2. Les lactoprotéines ___ 46

I.3.3. Les lipides du lait ___ 57

I.4. Technologie des puces à ADN « deux couleurs » pour l'étude du transcriptome ___ **71**

I.4.1. Définition et principe ___ 72

I.4.2. Les différents types de « puces à ADN » ___ 73

I.4.3. Les différentes étapes mises en œuvre pour la réalisation d'une analyse transcriptomique différentielle sur puces à ADN ___ 76

I.4.4. Analyse des données issues des puces à ADN ___ 83

II. Chapitre 2 : étude transcriptomique de la différenciation terminale du tissu mammaire caprin ___ ***90***

II.1. Projet de l'article 1 : La différenciation terminale des tissus mammaires caprins au cours de la gestation requière l'expression de gènes du système immunitaire ___ **91**

II.1.1. Introduction 92
II.1.2. Materials and methods 94
II.1.3. Results 102
II.1.4. Discussion 122
II.1.5. Conclusion 128
II.1.6. Acknowledgments 129

II.2. Comparaison directe 46 jours de gestation et 40 jours de lactation 129
II.2.1. Matériels et méthodes 129
II.2.2. Résultats 132
II.2.3. Discussion 139
II.2.4. Conclusion 141

III. Chapitre 3 : effet de la mutation K232A au locus DGAT1 sur les profils d'expression génique des tissus mammaires bovin 142

III.1. Facteurs de variation du taux butyreux et de la composition en acides gras du lait des ruminants 142
III.1.1. Les facteurs environnementaux modulant la matière grasse laitière 142
III.1.2. Les polymorphismes génétiques impactant la composition des laits bovins 144

III.2. Du QTL bovin affectant la quantité de matière grasse laitière à l'enzyme diacylglycérol acyltransferase 1 (DGAT1) 145
III.2.1. Du QTL affectant la quantité de matière grasse laitière au gène DGAT1 145
III.2.2. Du gène DGAT1 à l'enzyme fonctionnelle 147
III.2.3. Effet d'une invalidation ou d'une mutation sur le gène DGAT1 149

III.3. Effet du polymorphisme K232A au locus *DGAT1* sur l'activité transcriptomique de la glande mammaire bovine : Article 2 (en préparation) 150
III.3.1. Introduction 150
III.3.2. Matériels et méthodes 151
III.3.3. Résultats 161
III.3.4. Discussion 176
III.3.5. Conclusion 181

Discussion générale et conclusions 182

Bibliographie 192

Annexes 207

Abstract 234

Liste des abréviations

AA	Acide aminé
AAE	Acide aminé essentiel
AANE	Acide aminé non essentiel
ABCG2	« ATP-binding cassette, sub-family G, member 2 »
ACBP	« Acyl CoA Binding Protein »
ACC	Acétyl CoA Carboxylase
ACS	Acétyl CoA Synthétase
ADN	Acide Désoxyribonucléique
ADRP	« Adipocyte differentiation related protein » ou adipophiline
AG	Acide gras
AGPAT	Acylglycérol-3-phosphate acyltransférase
ARN	Acide RiboNucléique
ARNm	ARN messager
ARNpm	ARN pré-messager
ARNt	ARN de transfert
ATP	Adénosine Triphosphate
BSA	« Bovine Serum Albumine »
BASE	« BioArray Software Environment »
BTA	Chromosome bovin (Bos TAurus)
BTN	Butyrophiline
CD36	« Cluster of Differentiation » 36
CEM	Cellule Epithéliale Mammaire
CM	ChyloMicron
CNIEL	Centre National d'Interprofession et d'Economie Laitière
CLA	« Conjugated Linoleic Acid »
CRB-GADIE	Centre de Ressource Biologique – Génomique des Animaux Domestiques et d'Intérêt Economique
DGAT	DiacylGlycérol AcylTransférase
DHAP	DiHydroxyAcetone Phosphate
EGF	« Epidermal Growth Factor »
ERα	« Estrogen Receptor α »
EST	« Expressed Sequence Tag »
FABP	« Fatty Acid Binding Protein
FADS	« Fatty Acid DeSaturase »
FAS	« Fatty Acid Synthase »
FDR	« False Discovery Rate »
FGF	« Fibroblast Growth Factor »
FOI	« Frequence Of Incorporation »
FWER	« Family Wise Error Rate »
GEO	« Gene Omnibus »
GH	« Growth Hormone »
GK	Glycérol Kinase
GO	« Gene Ontology »
GPAT	Glycérol-3-Phosphate AcylTransférase
GPL	Génomique et Physiologie de la Lactation
GT	Galactosyltransférase

GTP	Guanosine TriPhosphate
IGF	« Insulin-like Growth Factor »
IPA	« Ingenuity Pathway Analysis »
KO	« Knock Out »
LDL	« Low Density Lipoprotein »
LPAAT	LysoPhosphatidique Acid Acyltransferase
LPL	LipoProtein Lipase
LPP	Acid Phosphatidique Phosphatase
MFGM	« Milk Fat Globule Membrane »
MGL	Matière Grasse Laitière
MIAME	« Minimum Information About a Microarray Experiment »
MP	Matière Protéique
MFP	« Mammary Fat Pad »
MUC1	Mucine 1
NADPH	Nicotinamide Adénine Dinucléotide Phosphate
NCBI	« National Center for Biotechnology Information »
PABP	« Poly(A) Binding Protein »
PCR	« Polymerase Chain Reaction »
PL	« Placental Lactogene »
PolII	ARN polymérase II
PRβ	« Progesterone Receptor β »
PTHrP	« ParaThyroid Hormone related Protein »
QTL	« Quantitative Trait Loci »
RE	Réticulum Endoplasmique
REG	Réticulum Endoplasmique Granuleux
RP-HPLC	« Reverse Phase-High Performance Liquid Chromatography »
SAGE	« Serial Analysis of Gene Expression »
SAM	« Significance Analysis Microarray »
SCD	Stéaroyl CoA Désaturase
SDS	« Sodium Dodecyl Sulfate »
SIGENAE	Système d'Information des GENomes des Animaux d'Elevage
SNP	« Single Nucleotide Polymorphism »
SnRNP	« Small nuclear RiboNucleoProtein »
SRP	« Signal Recognition Particle »
SSC	« Saline-Sodium Citrate »
SSR	« Signal Sequence Receptor »
TB	Taux Butyreux
TG	TriGlycéride ou triacylGlycérol
TGN	« Trans Golgi Network »
TL	Taux de Lactose
TLDA	« TaqMan® Low Density Array »
TP	Taux Protéique
UTR	« UnTranslated Region »
VLDL	« Very Low Density Lipoprotein »
WAAP	« World Association for Animal Production »
XDH	Xanthine (Oxidase) DeHydrogenase
XDR	« eXtended Dynamic Range »
αlac	α-lactalbumine (LALBA)

Liste des Tableaux

Tableau I-1 : Teneurs moyennes des principaux constituants du lait de différentes espèces[1] . 43
Tableau I-2 : Concentration et caractéristiques des protéines du lait de vache[1] 48
Tableau I-3 : Variants génétiques aux protéines majeures des principaux ruminants 51
Tableau I-4 : Composition de la matière grasse laitière et distribution des différentes classes de lipides[1] 58
Tableau I-5 : Caractéristiques des trois puces à ADN utilisées dans ce travail de thèse 74
Table II-1 : Primers used in this study. Each pair of primers amplifies the target cDNA (amplicon sizes ranging between 54-229 nucleotides). Primers pairs were designed with the Primer Express Software v2.0 (Applied Biosystems) excepts for 24S ribosomal protein primers which were manually designed. 101
Table II-2 : IPA Molecular and Cellular Function of gene sets enriched with most highly expressed genes at each developmental stage 103
Table II-3 : 82 probes with high signal intensity in each developmental stages..... 105
Table II-4 : Identification of the IPA biological functions associated to the differential analysis between two successive time points 108
Table II-5 : Top 25 of up-regulated and down-regulated transcripts in the goat mammary tissue at 40 days of lactation compared to 110 days of pregnancy (adjusted P-value<0.05). 111
Table II-6 : IPA biological functions associated to each cluster 114
Tableau II-7 : Liste des gènes différemment exprimés au niveau du tissu mammaire entre 46 jours de gestation et 40 jours de lactation. Le seuil de significativité est de 1% avec la correction des probabilités par la méthode de Benjamini-Hochberg. 134
Tableau III-1 : Amorces des systèmes optimisés pour doser les transcrits en PCRq..... 158
Tableau III-2 : Effets du génotype au locus *DGAT1* sur les performances zootechniques, la taille des globules gras et des micelles de caséines 162
Tableau III-3 : Effet du polymorphisme au locus DGAT1 sur la composition en acides gras des laits 163
Tableau III-4 : Phénotypage (variants protéiques) des animaux du dispositif pour la β-lactoglobuline et les caséines κ et β par analyse des laits écrémés en RP-HPLC..... 164
Tableau III-5 : Fonctions métaboliques déterminées par « Ingenuity Pathway Analysis » sur les gènes significativement surexprimés avec le génotype « faible » (GC/GC, A232) au locus *DGAT1* 169

Tableau III-6 : Fonctions métaboliques déterminées par « Ingenuity Pathway Analysis » sur les gènes significativement sous-exprimés avec le génotype « faible » (GC/GC, A232) au locus *DGAT1* 171
Tableau III-7 : Résultats sur la mise en relation des données transcriptomiques et phénotypiques..... 175

Liste des Figures

Figure I-1 : Schéma de la morphologie et de l'anatomie de la glande mammaire bovine en vue postérieure (adapté de Frandson, 1986; Quinn, 1980). 21
Figure I-2 : Schéma d'une cellule épithéliale mammaire sécrétrice et de sa membrane basale 23
Figure I-3 : Structure schématique d'un lobule mammaire (adapté de Larson, 1985; Whittemore, 1980) 24
Figure I-4 : Structure schématique d'un lobe mammaire (adapté de Quinn, 1980) 25
Figure I-5 : Coupe longitudinale schématique d'un trayon de glande mammaire bovine (adapté de Frandson, 1986; Quinn, 1980) 27
Figure I-6 : Développement de la glande mammaire bovine pendant la vie fœtale (Jammes & Djiane, 1988) 30
Figure I-7 : Mise en place des canaux lobulaires et du système lobulo-alvéolaire (Jammes & Djiane, 1988) 31
Figure I-8 : Coupes sagittales de la glande mammaire de génisse en première gestation (Swanson & Poffenbarger, 1979) 34
Figure I-9 : Représentation chronologique du développement et de la différenciation de la glande mammaire bovine au cours de la vie d'une femelle et régulations hormonales de ces étapes de développement (adapté de Jammes & Djiane, 1988; Neville *et al.*, 2002). 36
Figure I-10 : Représentation schématique des changements de concentration des hormones stéroïdiennes 38
Figure I-11 : Réaction de synthèse du lactose 43
Figure I-12 : Utilisation du glucose par la cellule épithéliale mammaire et voie de synthèse du lactose (d'après Bernard *et al.*, 2008; Guinard-Flament *et al.*, 2006; Kuhn *et al.*, 1980) 44
Figure I-13 : Origine du glucose capté par la cellule épithéliale mammaire (d'après Pougheon & Goursaud, 2001) 45
Figure I-14 : Teneur en protéines (g/l) du lait de neuf espèces (Martin *et al.*, 2003) 47
Figure I-15 : Quelques caractéristiques des caséines bovines (d'après Neveu, 2004; Swaisgood, 2003) 49
Figure I-16 : Organisation génomique du locus des caséines bovines (Martin *et al.*, 2002). .. 53
Figure I-17 : Organisation structurale des unités de transcription des caséines, de l'α-lactalbumine et de la β-lactoglobuline bovine (Martin *et al.*, 2002) 54

Figure I-18 : Représentation schématique d'un ARNm du tissu mammaire codant les protéines majeures du lait : les 4 caséines, la β-lactoglobuline et l'α-lactalbumine (Mepham *et al.*, 1992). 55
Figure I-19 : Origine des substrats utilisés par la cellule épithéliale mammaire (CEM) pour la synthèse des lipides du lait (d'après Doreau & Ferlay, 1994). 59
Figure I-20 : Synthèse des acides gras et des triglycérides dans la cellule épithéliale mammaire des ruminants (d'après Bernard *et al.*, 2008 ; Hawke & Taylor, 1995). 60
Figure I-21 : Désaturation de l'acide stéarique (C18:0) en acide oléique (C18:1) par la delta-9 désaturase 62
Figure I-22 : Synthèse des triglycérides dans la cellule épithéliale mammaire (adapté de Bell *et al.*, 1981; Ollier, 2007). 64
Figure I-23 : Enzymes intervenant dans la synthèse des phospholipides du lait (adapté de Bell *et al.*, 1981)..... 66
Figure I-24 : Formation de la gouttelette lipidique entre les deux couches phospholipidiques de la membrane du réticulum (Brown, 2001)..... 68
Figure I-25 : Nomenclature utilisée dans la schématisation des plans expérimentaux en technologie de puces à ADN « deux couleurs ». 77
Figure I-26 : Représentation schématique de différents types de plans expérimentaux (adapté de Kerr & Churchill, 2001). 78
Figure I-27 : Les différents types de marquages en technologie de puces à ADN « deux couleurs ». 81
Figure I-28 : Exemples de graphes réalisés dans la normalisation de données issues de la technologie de puces à ADN « deux couleurs ». 85
Figure II-1 : Loop experimental design and hybridization table 97
Figure II-2 : K-means clustering allocating into 19 clusters the 2,330 probes equivalent to 1,696 genes that varied significantly at least once 113
Figure II-3 : Gene expression profiles obtained by qPCR and microarrays. 120
Figure II-4 : Gene expression profiles obtained by qPCR and microarrays of key genes in clustering. 121
Figure II-5 : Gene expression profiles obtained by qPCR. 121
Figure II-6 : Comparaison des profils d'expression d'une vingtaine de gènes impliqués dans la synthèse et la sécrétion des lipides et protéines du lait ou impliqués dans la mise en place du tissu mammaire, obtenus par la technique d'analyse sur microarrays (barres blanches) et en PCRq (barres noires). 138

Figure III-1 : Du gène DGAT1 bovin à la protéine mature..... 147
Figure III-2 : Généalogie des animaux utilisés dans l'expérimentation..... 152
Figure III-3 : Analyse de la fraction « protéines majeures » des laits en RP-HPLC. 164
Figure III-4 : Analyse en composante principale sur les variables phénotypiques. 166
Figure III-5 : Comparaison des quantifications en PCR quantitative (histogramme blanc) et en microarray (histogramme noir) exprimées en moyenne des ratios de transcrits DGAT1 entre génotypes (GC, A232 / AA, K232). Les barres verticales représentent l'écart-type. 174

Contexte

La filière laitière française représente plus de 3 800 000 vaches laitières, 853 000 chèvres laitières et 1 483 000 brebis laitières réparties dans 94 300 exploitations employant directement 200 000 personnes (source : rapport annuel du CNIEL 2007 et site Internet de la « World Association for Animal Production », WAAP). La France est le deuxième producteur européen de lait après l'Allemagne. Environ 22 milliards de litres de lait sont collectés et livrés annuellement à plus de 700 entreprises de transformation qui emploient 60 000 personnes. L'exportation du lait et des produits laitiers représente un excédent commercial de 2,9 millions d'Euros. La filière laitière constitue par conséquent un pôle économique important pour la France.

L'augmentation de la consommation de lait (et des produits laitiers) dans les pays à forte croissance économique (Chine et Inde) et la reconversion à l'échelle mondiale des cultures d'élevage en cultures pour la production de bioénergies, ont entraîné, pour la première fois en 2007, une augmentation du prix du lait liée à une faible pénurie (source : rapport annuel du CNIEL 2007). Actuellement, en Europe, une restructuration de la filière est envisagée, en raison notamment d'une probable suppression des quotas laitiers.

Les attentes des différents acteurs de la filière laitières sont variées. Les producteurs cherchent à produire à moindre coût, la paie du lait doit leur permettre à la fois de couvrir les apports pour produire et de leur assurer un revenu. Les transformateurs sont soucieux d'un approvisionnement et d'une composition du lait stables et d'un prix le moins élevé possible. Les consommateurs, de plus en plus conscients de l'impact de l'alimentation sur leur santé, porteront attention à ce que le lait et les produits laitiers qu'ils consomment soient « naturellement » (source : satellite Bleu Blanc Cœur, 3R 2008, Paris, 3-4 décembre) sains, bons et bénéfiques pour leur santé. Enfin, la société, se rend compte de son impact sur son milieu de vie et de la nécessité de réagir en adoptant des techniques plus respectueuses de l'environnement, c'est le développement durable. De chacune de ces attentes, il ressort deux critères qu'il semble important de comprendre et maîtriser : la production laitière et la composition du lait.

Le lait est un produit complet constitué d'eau, de glucides, de protéines, de lipides et de minéraux, dont les effets des composants fins sur la transformation industrielle et la santé humaine sont de plus en plus connus. Par exemple, l'évolution des connaissances sur les différents acides gras a mis en évidence dans la viande, le lait et les produits laitiers des

ruminants la présence d'un acide gras reconnu pour son effet anti-cancérigène : l'acide ruménique (C18:2 *cis9, trans11*) ou « Conjugated Linoleic Acid » ou CLA (pour revue : (Williams, 2000). Une augmentation de la teneur de cet acide gras dans le lait diminue la fermeté des beurres et des fromages (Jones *et al.*, 2005).

De nombreux leviers existent pour modifier la productivité d'un animal ou la composition de son lait. Un premier moyen concerne les facteurs environnementaux ou les techniques d'élevage tels que la fréquence de traite, l'alimentation ou la gestion des génisses. Ainsi, une diminution de la fréquence de traite (passage de deux traites par jour, à une traite) entraîne une baisse de la production laitière et une augmentation des matières sécrétées (matière protéique et matière grasse). A l'inverse, une augmentation de la fréquence de traite augmente la production laitière mais n'a pas d'effet sur la matière (Lollivier *et al.*, 2002). L'alimentation a également un impact sur la productivité et la composition du lait. En comparaison à un ensilage de maïs, l'ensilage de foin ou d'herbe augmente la production laitière, diminue le taux butyreux (TB) et augmente la teneur en acides gras mono et poly-insaturés des laits (Chilliard *et al.*, 2000; Rego *et al.*, 2004). Les effets des supplémentations de l'alimentation, notamment l'apport lipidique dépendront de la nature de la supplémentation (tournesol, lin, colza), de la dose (3, 5% de la ration) et de la forme d'apport (huile, graine broyée, extrudée…). Le mode d'élevage des génisses peut impacter la productivité de l'animal. En effet, en période pré pubère le tissu adipeux de la glande mammaire se développe proportionnellement au parenchyme épithélial, ainsi une alimentation trop riche de l'animal à cette période peut entraîner une baisse de production laitière en lactation puisque le tissu adipeux se sera développé au détriment du parenchyme épithélial et donc du tissu sécrétoire (Jammes & Djiane, 1988).

Le deuxième moyen pour maîtriser la productivité d'un animal ou la composition de son lait est la génétique. Parmi les trois principales races bovines laitières françaises, la Prim'Holstein est la plus productive alors que la Normande présente les taux butyreux et protéique (TB et TP) les plus élevés (source : Institut de l'Elevage, FCL). Un bel exemple de l'effet de la génétique sur la composition du lait est celui du polymorphisme au locus *CSN1S1* spécifiant la caséine α_{S1} chez la chèvre. Il existe pour ce gène plusieurs allèles se répartissant en 4 catégories selon la teneur en caséine α_{S1} trouvée dans le lait : fort (3,6 g/l), intermédiaire (1,6 g/l), faible (0,45 g/l) ou nul (traces, Grosclaude *et al.*, 1994). Il est observé chez les individus porteurs des allèles faibles ou nuls, une baisse du TP, du TB et du rendement fromager.

Par ailleurs, l'émergence de la génomique, science qui étudie le génome, apporte une meilleure connaissance des génomes des espèces domestiques. Le découpage par des techniques de cartographie, du génome bovin en plusieurs marqueurs, permet la détection de nombreuses régions d'intérêt ou « Quantitative Trait Loci » (QTL) et pose les fondements d'un nouveau type de sélection des animaux : la sélection génomique (Bidanel *et al.*, 2008). L'affinement dans la connaissance du génome permet parfois de détecter les mutations génétiques expliquant un QTL. Ainsi, les QTL affectant la matière grasse laitière ont pu être associés à une mutation non-sens au niveau du gène ABCG2 pour le QTL sur le chromosome bovin 6 (BTA 6, Cohen-Zinder *et al.*, 2005; Olsen *et al.*, 2007), une double mutation au niveau de l'exon 8 du gènes DGAT1 pour le QTL du BTA 14 (Grisart *et al.*, 2002; Winter *et al.*, 2002) et une mutation dans le gène SCD pour le QTL du BTA 26 (Taniguchi *et al.*, 2004).

En revanche, à notre connaissance, il n'a pas été encore décrit dans la littérature comment le déterminisme génétique (mutation causale entre autres) se répercute au niveau du fonctionnement des tissus mammaires.

C'est dans ce contexte que mes travaux de thèse ont débuté. Leurs objectifs étaient d'identifier les gènes ou réseaux de gènes impliqués dans la différenciation terminale des tissus mammaires caprins et dans la variation de TB et de composition en acides gras (AG) du lait induite par la mutation K232A au locus *DGAT1* chez le bovin. La stratégie envisagée au début de ce travail était l'utilsation d'une approche globale en étudiant simultanément les variations d'expression de milliers de gènes dans différents états physiologiques de la glande mammaire.

Cette thèse s'articule en plusieurs parties comprenant, pour introduire le sujet, une synthèse bibliographique décrivant : (i) la glande mammaire et son développement ; (ii) la composition du lait, la synthèse et la sécrétion de ses constituants majeurs ; (iii) une analyse succincte des techniques d'étude du transcriptome, en décrivant de façon plus précise la technique des puces à ADN deux couleurs, utilisée dans ce travail de thèse. Cette première partie, assez étoffée, établit les fondations du travail réalisé dans cette thèse. En effet, une bonne connaissance des mécanismes de mise en place et de fonctionnement de la glande mammaire est la base pour identifier et mieux caractériser les gènes intervenant dans ces mécanismes. La deuxième partie détaille l'étude transcriptomique de la différenciation terminale de la glande mammaire caprine au cours de la gestation. Finalement, la troisième partie analyse l'effet du polymorphisme au locus *DGAT1* sur les performances laitières, la

composition et les caractéristiques des structures supramoléculaires du lait en liaison avec l'activité transcriptionnelle des tissus de la glande mammaire bovine.

I. Chapitre 1 : Synthèse bibliographique

I.1. Description morphologique et anatomique de la glande mammaire des ruminants

Les descriptions morphologiques et anatomiques de la glande mammaire présentées dans cette partie, sont adaptées de A.W. Lishman, « The cow's udder and milk secretion », Department of Animal Science University of Natal (http://agriculture.kzntl.gov.za/portal/publications/production_guidelines/dairying_in_natal/dairy6_1.htm, 2008) ;
et P. Lacasse, cours sur la Biologie de la Lactation, Département de Biologie, Université de Sherbrooke, (http://pages.usherbrooke.ca/infosbio/PSL705/Biologie/course-f.htm, 2008).

La mamelle est une glande exocrine composée de quatre (bovins) ou deux (ovins et caprins) quartiers indépendants, située sur la face ventrale de l'animal en position inguinale. Les quartiers de droite et de gauche de la mamelle sont séparés par un ligament de suspension central composé de tissu élastique. Des branches de ce ligament peuvent s'étendre dans les quartiers. La mamelle est recouverte d'une peau élastique. Elle peut s'agrandir sous l'effet de l'accumulation du lait entre deux traites ou deux tétées. Dans le cas où le ligament central est faible, la mamelle pend trop, ce qui peut entraîner des difficultés pour la traite et une exposition plus importante à de probables agents pathogènes due au rapprochement des trayons avec le sol. Chez la vache, les quartiers antérieurs et postérieurs sont séparés par une fine membrane composée de tissu conjonctif. Il est possible d'observer les veines et vaisseaux sanguins sous-cutanés qui irriguent la mamelle.
En lactation, chaque quartier contient un tissu sécrétoire constitué de cellules épithéliales mammaires, des canaux galactophores, une citerne de la glande et un trayon (cf. Figure I.1).

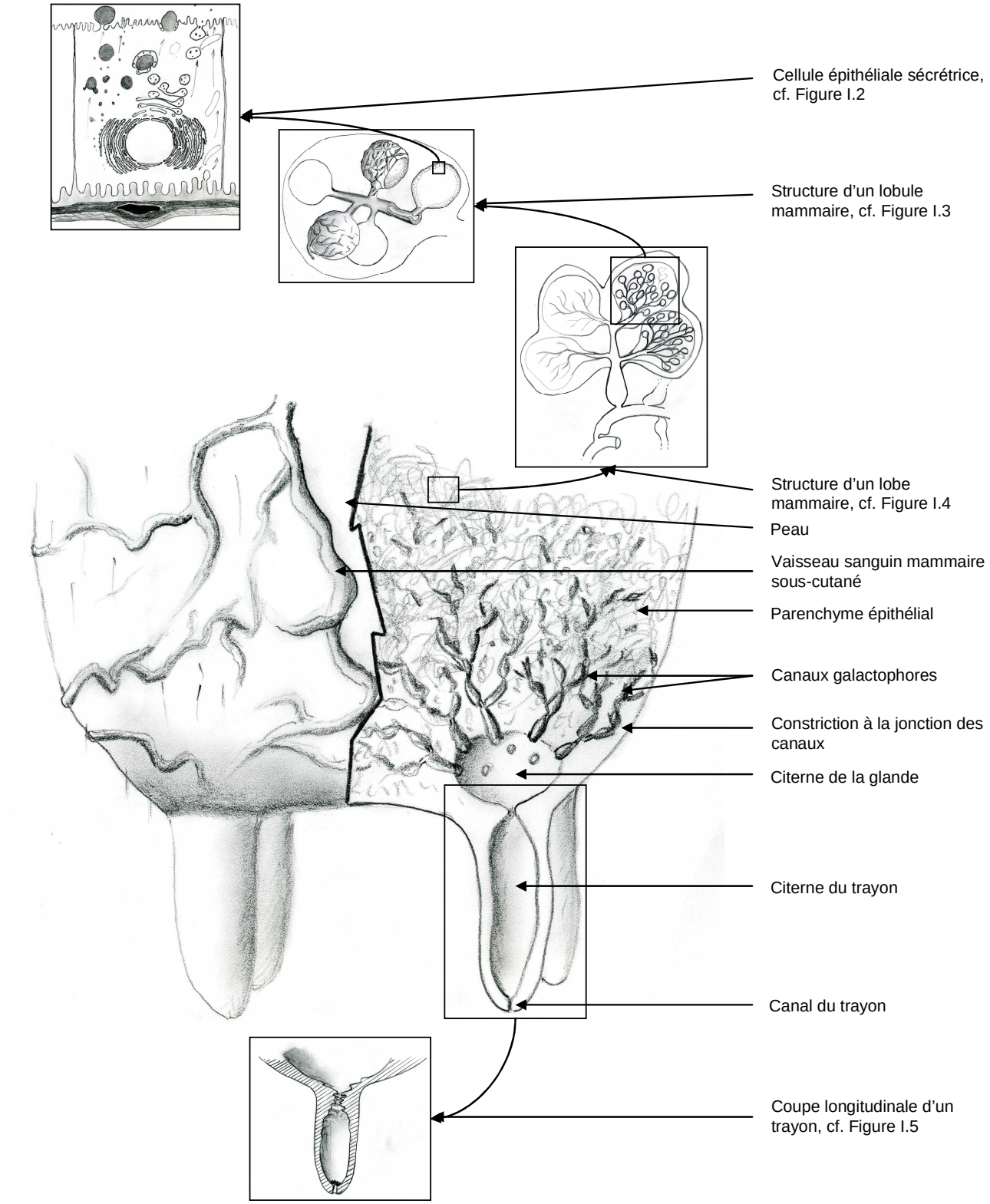


Figure I-1 : Schéma de la morphologie et de l'anatomie de la glande mammaire bovine en vue postérieure (adapté de Frandson, 1986; Quinn, 1980).
Sur la partie gauche, la morphologie externe de la glande mammaire. Sur la partie droite, une coupe longitudinale de la glande mammaire dans l'axe du milieu du trayon. Chaque zoom fait référence à un schéma légendé dans les figures suivantes

I.1.1. L'unité de production du lait : la cellule épithéliale mammaire

La cellule épithéliale mammaire (CEM) est une cellule sécrétrice (Figure I.2) constituant la plus petite unité des alvéoles (ou acini). En lactation, les CEM sont polarisées, avec la face basale située du côté de la membrane basale et la face luminale située du côté de la lumière alvéolaire. Les constituants du lait sont sécrétés dans la lumière alvéolaire par la face luminale. Les CEM sont liées entre elles par des jonctions serrées et reposent sur une membrane basale constituée de laminine, de collagène et de glycosaminoglycanes. Les CEM contiennent en partie basale le noyau entouré du réticulum endoplasmique granuleux (REG). En direction de la membrane plasmique apicale, le cytoplasme contient l'appareil de Golgi et les différentes unités de sécrétion : les gouttelettes lipidiques et les vésicules de sécrétion. La membrane plasmique apicale forme des microvillis (pour revue, cf. Mather & Keenan, 1998a). La synthèse des lipides du lait s'opère au sein du réticulum endoplasmique et se matérialise par la formation des gouttelettes lipidiques entre les deux couches membranaires du réticulum. Une fois formées, les gouttelettes lipidiques migrent vers la membrane apicale selon des mécanismes non encore élucidés. Le lactose est synthétisé au sein de l'appareil de Golgi et s'accumule dans des vésicules de sécrétion. Les protéines sont synthétisées par des ribosomes situés à la surface du REG. Elles passent ensuite dans l'appareil de Golgi ou débute le processus de maturation (phosphorylation, notamment) avant d'être incluses dans des vésicules de sécrétion.

La gouttelette lipidique est secrétée dans la lumière de l'acinus par enroulement progressif dans la membrane plasmique apicale (voie A et B sur la Figure I.2, revue par Bauman *et al.*, 2006). Lorsque la gouttelette est totalement entourée de membrane plasmique, elle est expulsée dans la lumière alvéolaire. Pendant leur migration dans la CEM, certaines gouttelettes lipidiques peuvent fusionner avec des vésicules de sécrétion, formant ainsi, des vacuoles. Leur excrétion se ferra conjointement aux vésicules de sécrétion (voie C sur la Figure I.2, Heid & Keenan, 2005). Par ailleurs, les vésicules de sécrétion non liées aux globules gras, libèrent leur contenu dans la lumière de l'acinus par exocytose, c'est à dire fusion membranaire (voie D sur la Figure I.2, Mather & Keenan, 1998b). Certaines protéines provenant du sang telles que les immunoglobulines, la sérum albumine, peuvent traverser la CEM par un mécanisme de transocytose (voie E sur la Figure I.2). Ces protéines sont emmagasinées dans une vésicule de sécrétion au niveau de la membrane basale puis excrétées dans la lumière de l'acinus par exocytose (Mather & Keenan, 1998a).

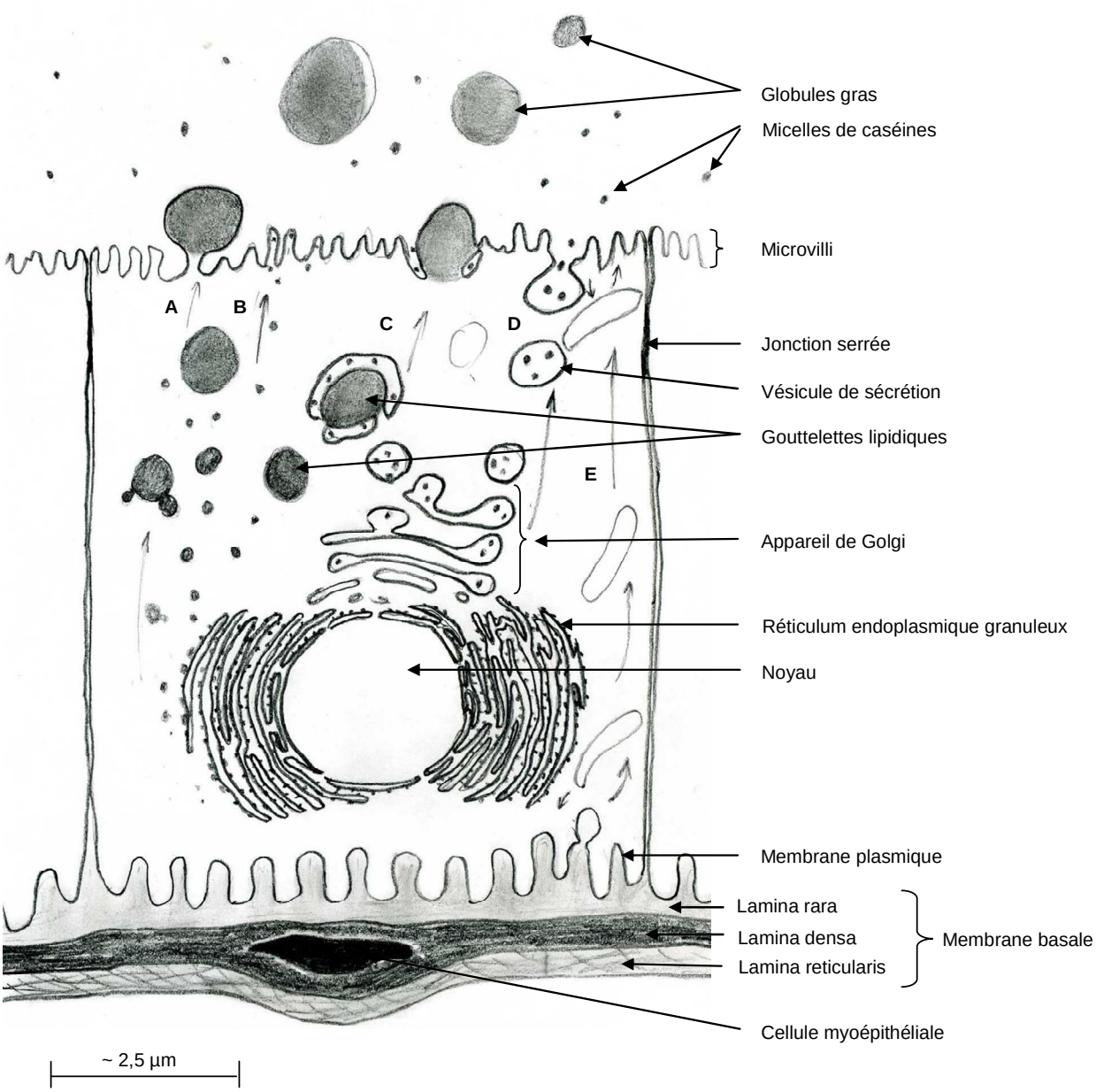


Figure I-2 : Schéma d'une cellule épithéliale mammaire sécrétrice et de sa membrane basale
Les voies de sécrétion des gouttelettes lipidiques (A et B), la voie de sécrétion conjointe des gouttelettes lipidiques et du contenu des vésicules de sécrétion (C), la voie de sécrétion du contenu des vésicules de sécrétion (D), et le mécanisme de transocytose (E) sont représentés ici (adapté de (Bauman *et al.*, 2006; Mather & Keenan, 1998b; Mather & Keenan, 1998a)

I.1.2. Le tissu sécrétoire

Les cellules épithéliales mammaires sont regroupées entre elles pour former l'unité de production du lait : l'alvéole. C'est une structure microscopique de forme presque sphérique dont la surface interne est tapissée d'un alignement d'une couche unique de cellules épithéliales sécrétrices (Figure I.3) et dont le centre, la lumière alvéolaire, est gorgé de lait pendant la lactation. Chaque alvéole est entouré de cellules myoépithéliales aidant à leur contraction pour l'éjection du lait vers les canaux galactophores. Des vaisseaux sanguins sont

également en contact avec les alvéoles, permettant ainsi l'approvisionnement des cellules épithéliales mammaires en nutriments et oxygène et leur régulation par des hormones.

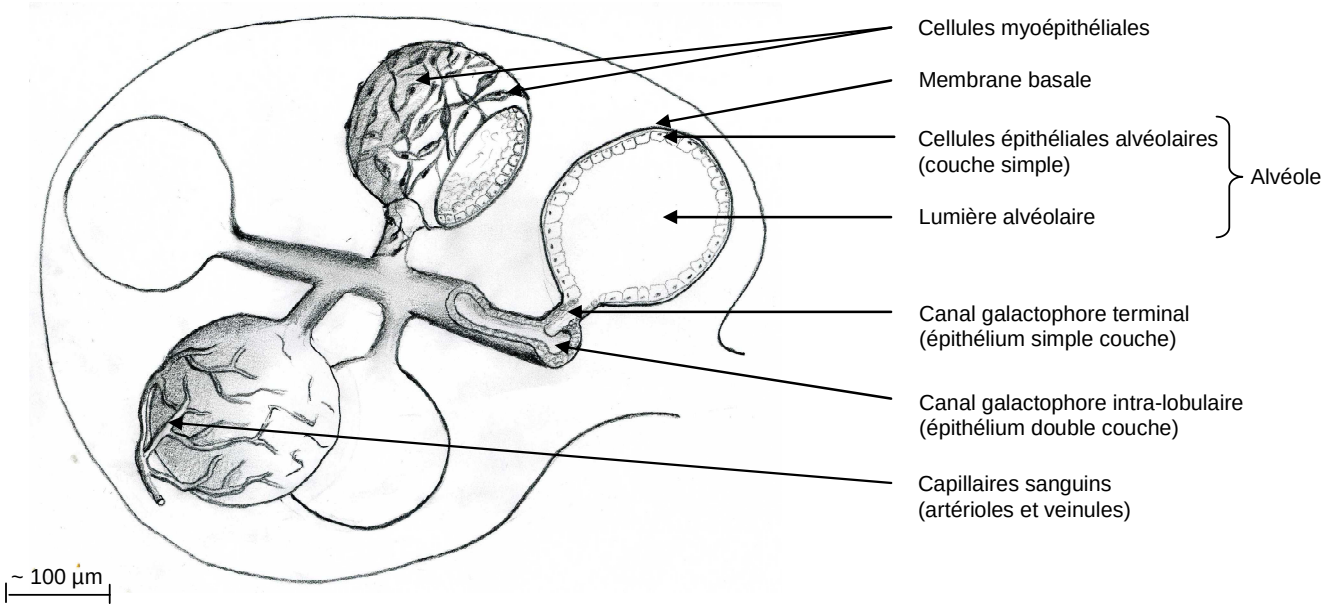


Figure I-3 : Structure schématique d'un lobule mammaire (adapté de Larson, 1985; Whittemore, 1980)

Un groupe d'alvéoles noyés dans des faisceaux conjonctifs forme un lobule. Ces lobules sont regroupés en lobes (Figure I.4), séparés par du tissu conjonctif. La mamelle est ainsi formée d'un ensemble de lobes glandulaires connectés par les canaux galactophores, constituant ainsi le parenchyme épithélial, et de tissu conjonctif fibreux et élastique composant pour partie le stroma.

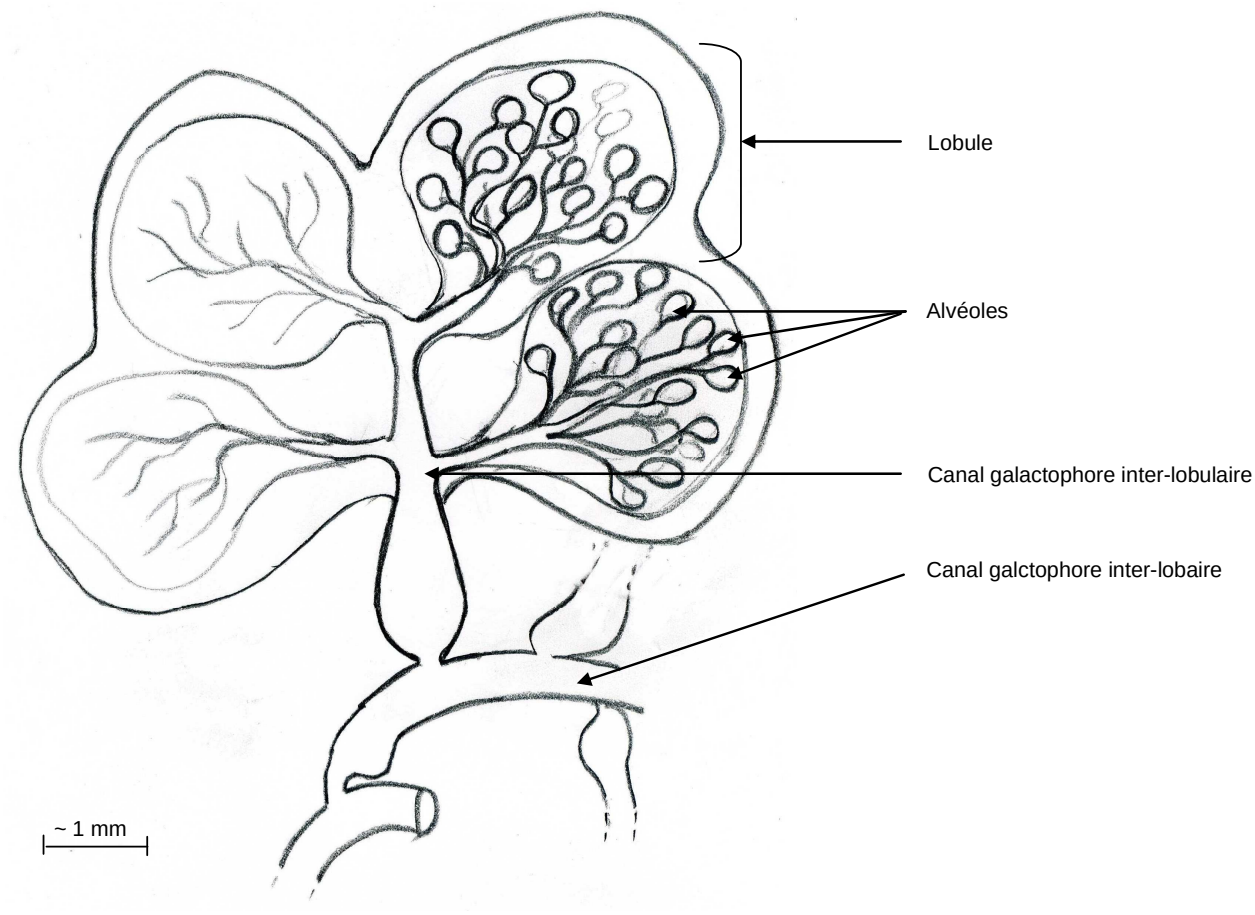


Figure I-4 : Structure schématique d'un lobe mammaire (adapté de Quinn, 1980)

I.1.3. Les canaux galactophores

Les petits canaux galactophores qui drainent chaque alvéole se rejoignent pour former des canaux tertiaires. Ces derniers se rassemblent en canaux secondaires puis primaires qui aboutissent à la citerne de la glande. Ce système de canaux a pour fonction de collecter le lait produit par l'épithélium sécrétoire, de conserver une partie de ce lait entre les traites et de transporter le lait jusqu'à la citerne de la glande. A chaque embranchement entre les canaux, il y a une constriction qui permet de retenir le lait jusqu'à ce que l'animal soit stimulé pour la sécrétion. Les canaux galactophores sont bordés par un épithélium à deux assises de cellules cylindriques (Figure I.3), mais dans les canaux les plus fins tels que ceux conduisant directement aux alvéoles, il n'existe qu'une seule assise de cellules épithéliales. Des cellules myoépithéliales entourent l'épithélium des canaux et des alvéoles et se contractent sous l'action de l'ocytocine, lors de l'éjection du lait.

I.1.4. La citerne de la glande

Il existe une citerne de la glande par quartier mammaire. La citerne de la glande mammaire (Figure I.1) correspond à de larges dilatations des canaux galactophores en sinus et en poches. Cette structure est particulièrement nette chez la chèvre et la brebis. Chez la vache, le volume de la citerne est de 400 à 500 ml. Au sein d'une même espèce, le volume de la citerne varie d'une race à l'autre, entraînant des capacités variables de stockage du lait et donc des temps différents entre les traites.

I.1.5. Les trayons

A la base de la citerne de la glande se trouve le trayon qui est le bout du pis, ou quartier mammaire, par lequel le lait est éjecté. Le trayon représente le premier contact ouvert entre le milieu extérieur et l'intérieur de la glande, il constitue ainsi la première protection de la glande face aux agents pathogènes.

Le trayon est séparé de la citerne de la glande par des plis annulaires de tissu constituant une barrière contre une invasion par des agents pathogènes. Le lait passe de la citerne de la glande dans une petite cavité du trayon, appelée citerne du trayon. Cette petite citerne emmagasine le lait drainé à partir de la glande. Elle peut contenir de 15 à 40 ml de lait, selon la taille du trayon. Le lait est ensuite évacué de la glande par un canal de 8 à 12 mm de long (Figure I.5). Ce canal est fermé entre deux traites par un muscle circulaire lisse appelé sphincter situé à son extrémité extérieure. Ce sphincter sert à la fois à garder le lait dans la glande, mais aussi à protéger celle-ci d'une invasion bactérienne. Les caractéristiques du sphincter et du canal sont importantes pour les critères productifs de l'animal. En effet, si le canal est petit ou si le sphincter ne se relâche pas bien, l'animal est difficile à traire et le temps de traite est plus long. En revanche, si le canal est large et le sphincter trop lâche, le lait sera évacué de la glande entre les traites et la glande restera ouverte au milieu extérieur. Les trayons contiennent de nombreux vaisseaux sanguins avec des petites valves qui maintiennent la circulation sanguine lorsque le trayon est massé. Un mauvais massage pendant la traite entraîne une rétention du sang dans le trayon, douloureuse pour l'animal.

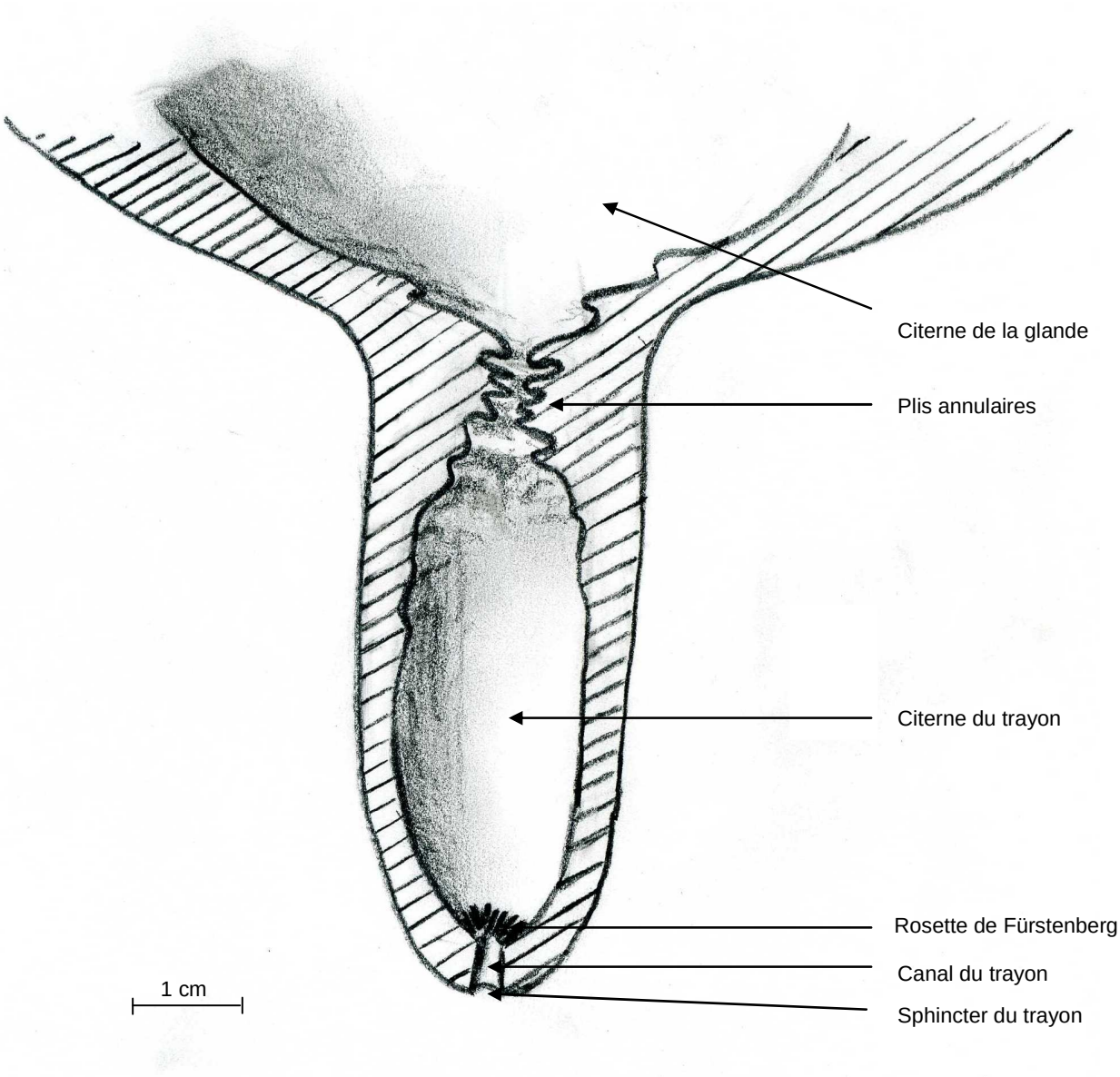


Figure I-5 : Coupe longitudinale schématique d'un trayon de glande mammaire bovine (adapté de Frandson, 1986; Quinn, 1980)

D'un point de vue histologique, le canal du trayon est tapissé d'un épithélium stratifié squameux qui forme 4 à 8 replis longitudinaux. A la jonction entre le canal et la citerne du trayon, ces plis s'élargissent pour former la rosette de Fürstenberg. Les cellules de cette structure produisent de la kératine pour former un film protecteur contenant des acides gras à longue chaîne ayant des effets bactériostatiques. Les replis de cette structure renferment également une concentration importante de lymphocytes (Rainard & Poutrel, 1993). La rosette de Fürstenberg constitue ainsi une première protection de la glande mammaire. La paroi du trayon est riche en fibres musculaires lisses, en fibres de collagène, en terminaisons nerveuses et en vaisseaux sanguins.

I.1.6. Les mécanismes d'évacuation du lait

Le lait est synthétisé dans les cellules épithéliales mammaires tapissant les alvéoles à partir des nutriments apportés par les vaisseaux sanguins qui viennent à leur contact. Le lait synthétisé est secrété dans la lumière alvéolaire. Il existe deux mécanismes d'évacuation du lait, le premier est l'écoulement du lait par contraction de muscles lisses, le deuxième correspond à un réflexe d'éjection.

Le premier mécanisme d'évacuation du lait est l'écoulement de celui-ci après ouverture du sphincter sous l'effet de la pression du trayon en début de traite. Ce mécanisme débute 5 à 10 secondes après la stimulation du trayon. Il implique la contraction des muscles lisses entourant les canaux, ce qui provoque l'évacuation du lait qu'ils contiennent. Cette phase d'écoulement permet l'évacuation de 40 à 50 % du lait produit.

Le second mécanisme est le réflexe d'éjection du lait. Lors des premières stimulations du trayon ou de l'animal (odeur, vision, ouie), des impulsions nerveuses vont du trayon (ou de tout autre organe sensoriel) jusqu'au cerveau qui libère alors dans le sang une hormone du complexe hypothalamo-hypophysaire : l'ocytocine. Cette hormone agit sur les cellules myoépithéliales entourant les alvéoles en provoquant leur contraction. Sous la contraction des cellules myoépithéliales, les alvéoles sont pressées entraînant l'éjection du lait contenu dans la lumière des acini vers les canaux terminaux puis les canaux intra lobulaires, inter lobulaires et inter lobaires où il atteint la citerne de la glande puis celle du trayon. Le réflexe d'éjection du lait a lieu habituellement 20 à 30 secondes après la stimulation initiale et termine environ 6 minutes après (Labussière, 1993).

I.2. Développement et différenciation de la glande mammaire des ruminants

I.2.1. Descriptions morphologiques du développement de la glande mammaire des ruminants

Les différentes étapes caractérisant l'acquisition des structures et fonctions de la glande mammaire sont la mammogenèse, la lactogenèse et la galactopoïèse.

La mammogenèse est l'étape correspondant à la croissance de la glande mammaire. Elle est caractérisée par le développement des canaux, leur ramification et l'apparition et le développement du tissu lobulo-alvéolaire. Ce processus s'étend de la vie embryonnaire à la première mise bas de l'individu. Il est discontinu au cours de la vie d'une femelle.

La lactogenèse est l'étape de différenciation cellulaire permettant l'acquisition d'une activité de synthèse et de sécrétion des cellules épithéliales, responsables de la production de lait. Cette étape se situe juste avant la parturition.

La galactopoïèse est la phase de sécrétion lactée dont la mise en place intervient à la parturition. Elle est entretenue par le tirage du lait (soit la traite, soit la tétée). La production décline toutefois au cours du temps et un arrêt de la traite ou le sevrage entraîne une régression du tissu lobulo-alvéolaire, appelée involution (revue par Jammes & Djiane, 1988).

Contrairement aux rongeurs dont l'épithélium mammaire se développe au contact d'un tissu adipeux (coussin adipeux mammaire, ou « Mammary Fat Pad », MFP), l'épithélium mammaire des humains et des ruminants se développe au contact d'un stroma plus fibreux essentiellement constitué de collagène et de fibroblastes (Hovey *et al.*, 2002; Sheffield, 1988).

I.2.1.1. Développement de la glande mammaire durant la vie fœtale

Chez le bovin, dès le 32ème jour de vie fœtale, des rudiments de quatre structures mammaires sont visibles sur la face ventrale de l'embryon, sous la forme d'un petit épaississement (cf. Figure I.6). Entre le 32ème et le 80ème jour de vie fœtale, la prolifération cellulaire donne naissance à un cordon de cellules appelé canal primaire. Ce canal s'arborise rapidement en canaux secondaires qui seront les futurs canaux lobulaires. La différenciation sexuelle s'effectue au stade du bourgeon mammaire (entre 50 et 80 jours de vie fœtale), chez les mâles une décharge de testostérone produite par les testicules fœtaux provoque une dégénérescence des cellules des canaux isolant le bourgeon mammaire et inhibant ainsi la formation ultérieure du mamelon (Jammes & Djiane, 1988). Au 80ème jour de vie fœtale, les vaisseaux sanguins et les tissus conjonctifs et adipeux constituant le stroma se forment à partir du mésoderme (Hennighausen & Robinson, 2001; Hovey *et al.*, 2002; Jammes & Djiane, 1988). Entre le 80ème et le 110ème jour de vie fœtale, la partie distale du canal primaire se creuse en lumière, formant l'ébauche de la citerne. Ces développements concernent principalement le parenchyme épithélial qui dérive de l'ectoderme.

Le mécanisme de développement de la glande mammaire au stade foetal chez les ovins est semblable aux mécanismes décrits ci-dessus (revue Hovey *et al.*, 2002). En revanche, il y a peu (voire pas) de données dans la littérature, chez les caprins.

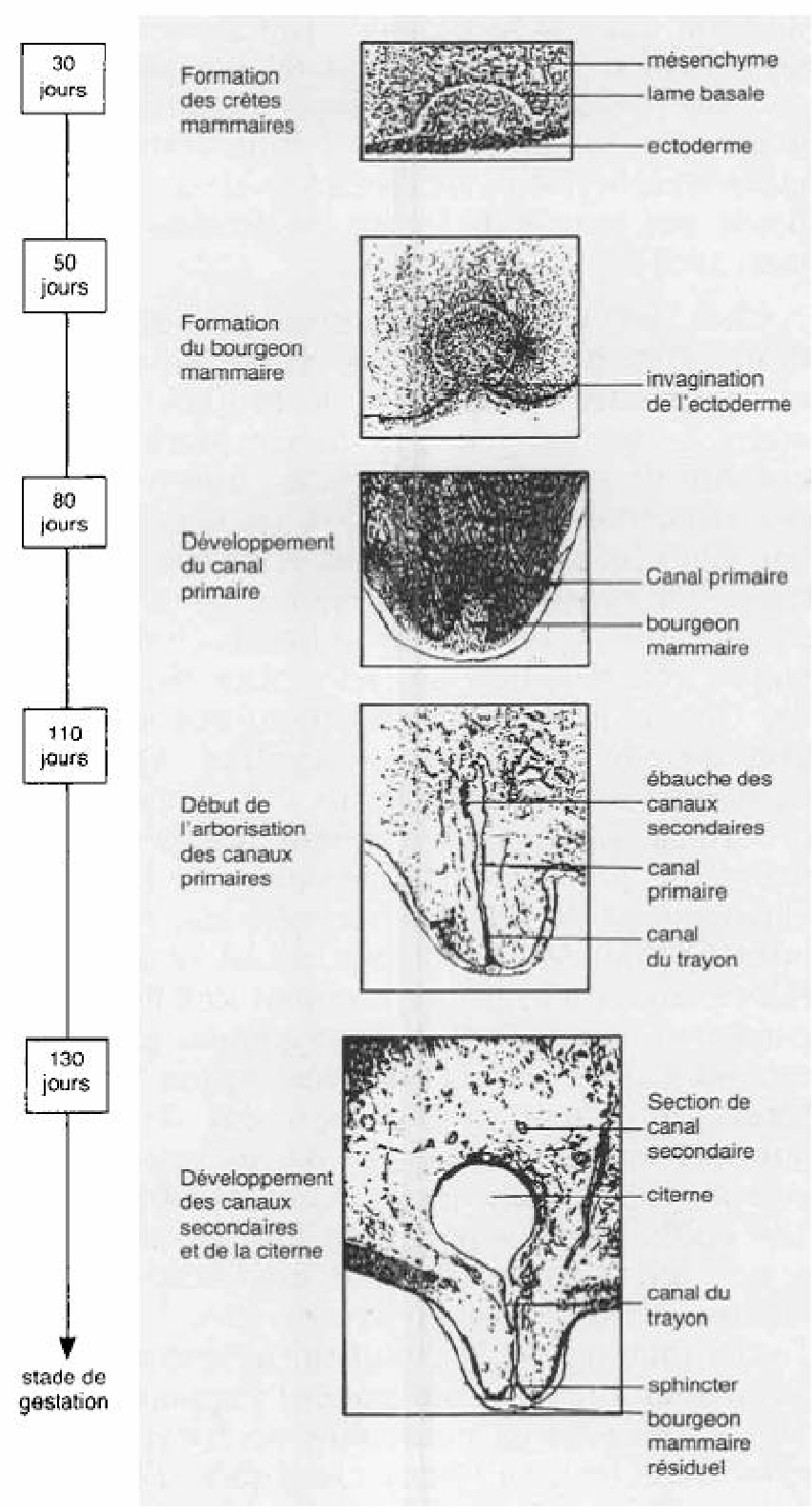


Figure I-6 : Développement de la glande mammaire bovine pendant la vie fœtale (Jammes & Djiane, 1988)

A la naissance, le mamelon est constitué d'un trayon formé d'un canal et d'un sphincter, d'une citerne d'où partent des canaux primaires, qui se ramifient en canaux

secondaires. Les canaux secondaires se terminent par de modestes lobules ramifiés et positionnés à proximité du stroma (revue de Hovey *et al.*, 2002).

I.2.1.2. Développement de la glande mammaire entre la naissance et la puberté

Après la naissance, le parenchyme mammaire de la génisse croît de façon isométrique, c'est-à-dire dans les mêmes proportions, avec le reste du corps. Trois mois environ après la naissance, commence, chez les ruminants, la période pré-pubère marquée par une croissance allométrique positive de la glande mammaire, c'est-à-dire que la glande croît plus vite que le reste du corps. Cette croissance se prolonge jusqu'après l'apparition des cycles sexuels chez la génisse, c'est-à-dire jusqu'à l'âge de 9 mois.

La phase de croissance pré-pubertale est caractérisée chez les ruminants par la prolifération des lobules terminaux, sans localisation précise de sites de croissance. A priori les lieux de division cellulaire, mitose, se situeraient à la périphérie du parenchyme (Hovey *et al.*, 2002). Pendant la croissance allométrique, la prolifération du tissu adipeux est conjointe à la croissance rapide des canaux lobulaires (Figure I.7). L'alimentation de la génisse est surveillée à cette période car une croissance trop rapide de la glande mammaire est associée à une diminution de la productivité laitière en lactation, due à un développement trop important du tissu adipeux, au détriment du parenchyme épithélial (revues d'Ellis, 1998; Jammes & Djiane, 1988).

Après la phase de croissance allométrique, la mamelle est à nouveau dans une phase de croissance isométrique (Sinha & Tucker, 1969).

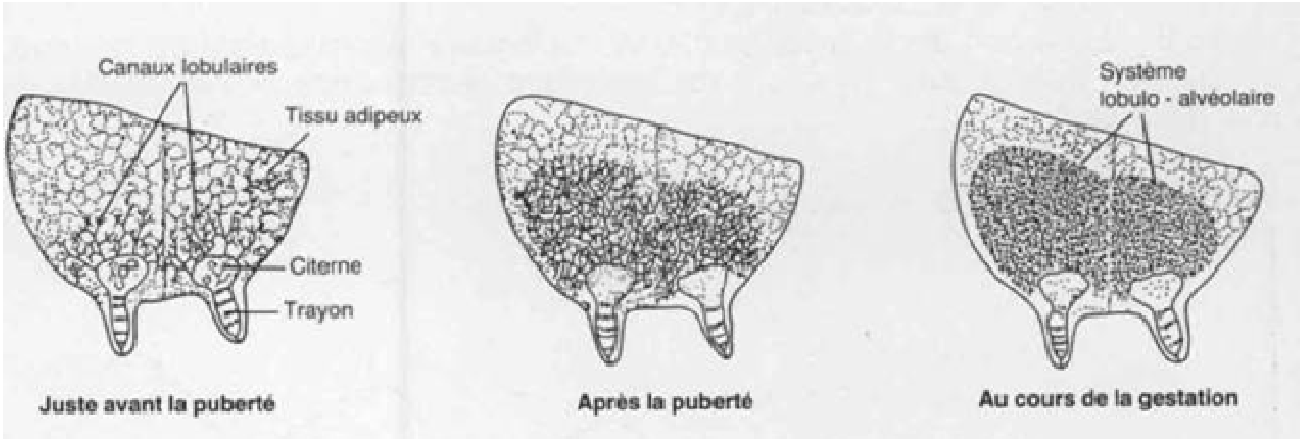


Figure I-7 : Mise en place des canaux lobulaires et du système lobulo-alvéolaire (Jammes & Djiane, 1988)

I.2.1.3. Développement de la puberté à la gestation, les cycles sexuels

La puberté, chez les ruminants, débute au moment de l'apparition des cycles oestraux, entre 6 et 9 mois après la naissance chez la génisse (Hovey *et al.*, 2002). A partir de ce moment, la glande mammaire est de nouveau dans une phase de croissance isométrique, marquée par des phases de développement/différenciation de l'épithélium sous l'action des hormones stéroïdiennes, en rapport avec les cycles sexuels.

Pendant la phase folliculaire du cycle sexuel, il y a développement des canaux mammaires sous l'action des oestrogènes. Pendant la phase lutéale, il y a développement des structures lobulo-alvéolaires sous l'action combinée des estrogènes et de la progestérone, bien que le développement lobulo-alvéolaire total ne se réalise qu'en gestation (Hovey *et al.*, 2000). La croissance mammaire est minime avant la première gestation, puisqu'elle n'exige pas la préparation finale du tissu pour la production de lait (P. Lacasse, cours sur la Biologie de la Lactation, Département de Biologie, Université de Sherbrooke, http://pages.usherbrooke.ca/infosbio/PSL705/Biologie/course-f.htm, 2008).

I.2.1.4. Développement et différenciation de la glande mammaire pendant la gestation

Il y a très peu de descriptions récentes dans la littérature du développement de la glande mammaire pendant la gestation chez les ruminants.

En revanche, les mécanismes de développement et de différenciation de la glande mammaire chez la souris sont bien décrits et se décomposent en trois étapes principales. Le premier mécanisme consiste en un développement des canaux pendant le premier tiers de la période de gestation. Ce mécanisme est suivi du développement des bourgeons terminaux pour former les structures alvéolaires. Enfin, la dernière moitié du temps de gestation consiste en un développement de ces structures par prolifération cellulaire et une différenciation des cellules alvéolaires en cellules épithéliales sécrétrices (Hennighausen & Robinson, 1998; Pitelka *et al.*, 1973).

Chez les ruminants, il est connu qu'après la puberté, le parenchyme épithélial est constitué de canaux primaires qui se ramifient en canaux secondaires, puis tertiaires et en de nombreux lobules. (Akers *et al.*, 2000) comparent visuellement les tissus mammaires des rongeurs et des ruminants. Ils rapprochent la structure du tissu mammaire des rongeurs à celle d'un arbre et celle des ruminants à celle d'un chou fleur. Le tissu mammaire des ruminants

n'a pas de structures en bourgeons terminaux comme chez les rongeurs (Ellis, 1998). De par les différences morphologiques entre les rongeurs et les ruminants après la puberté, il est supposé que les mécanismes morphologiques de développement sont différents entre ces deux groupes.

Chez la génisse en première gestation, une étude du développement morphologique (Swanson & Poffenbarger, 1979) décrit le développement du tissu mammaire, comme un envahissement progressif de tout l'espace mammaire composé de stroma et de tissu adipeux (cf. Figure I.8). Ces auteurs observent, qu'en début de gestation, la structure épithéliale croissante est constituée de nombreux tubules regroupés comportant une double couche de cellules épithéliales. A environ 5 mois de gestation (soit un peu plus de la moitié de la gestation) chez la vache, le tissu montre à la fois un épithélium avec une double couche de cellules, représentant les canaux, et un épithélium constitué d'une unique couche de cellules épithéliales, évoquant les alvéoles puisque ces structures contiennent du matériel sécrétoire. A 6-7 mois de gestation, le nombre et la taille des alvéoles augmentent. Leur contenu est composé de globules gras ainsi que d'un matériel supposé être des protéines. A 8 mois de gestation, les alvéoles se sont élargies et sont remplies de sécrétions colostrales contenant des globules gras. A 9 mois de gestation (peu avant la mise-bas), plusieurs alvéoles des glandes non traites sont semblables à celles des glandes mammaires en gestation, car elles contiennent à la fois des granules de sécrétion (protéines) et des globules gras plus larges et plus nombreux qu'aux stades de gestation précédents. Vers 8-9 mois de gestation, il reste seulement une fine couche de tissu adipeux à la base de la glande mammaire et aux extrémités avant et arrière de la glande, le tissu mammaire occupe toute la glande.

Chez les ruminants, le début de la gestation est ainsi marqué par la prolifération des lobules terminant les canaux secondaires et tertiaires. Dès la mi-gestation, les structures alvéolaires seraient formées. Au cours du dernier tiers de la gestation, s'ensuivraient la prolifération et la différenciation des cellules qui tapissent les alvéoles en cellules épithéliales sécrétrices. Les alvéoles seraient aussi entourées, au fur et à mesure, de cellules myoépithéliales formant alors de petits paniers (Hovey *et al.*, 1999).

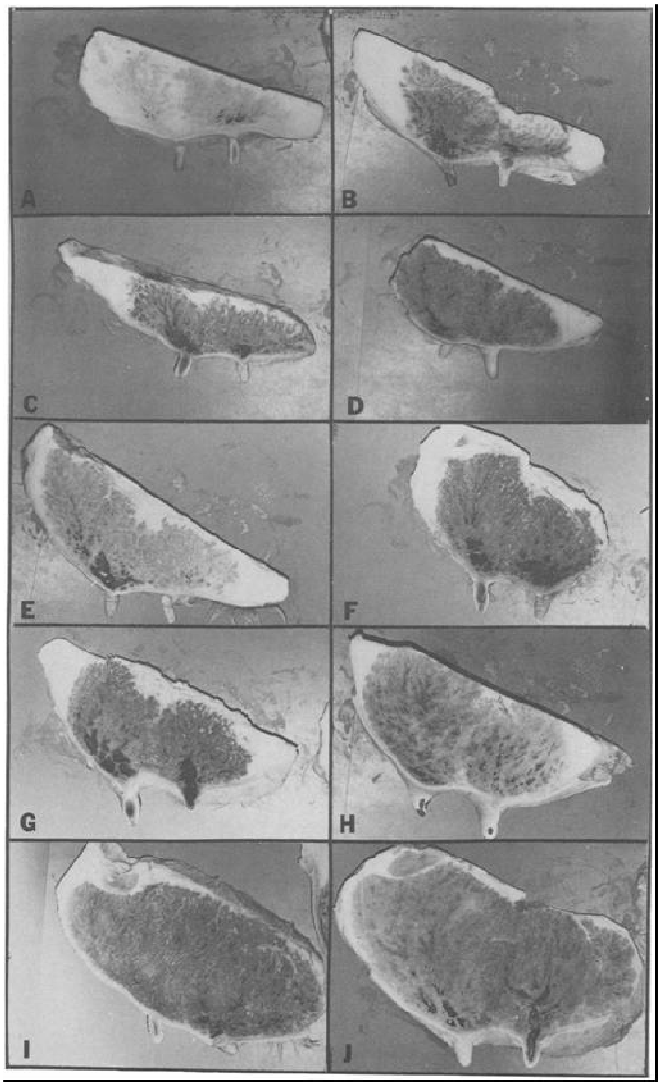


Figure I-8 : Coupes sagittales de la glande mammaire de génisse en première gestation (Swanson & Poffenbarger, 1979)
A : 0 mois de gestation ; B : 1 mois de gestation ; C : 2 mois de gestation ; D : 3 mois de gestation ; E : 4 mois de gestation ; F : 5 mois de gestation ; G : 6 mois de gestation ; H : 7 mois de gestation ; I : 8 mois de gestation ; J : 9 mois de gestation.

En fin de gestation, juste avant la parturition, le parenchyme mammaire se compose d'un épithélium sécrétoire tapissant les alvéoles, de cellules myoépithéliales et d'un épithélium formant les canaux. Le stroma environnant est composé de protéines édifiant une matrice incluant le collagène et la laminine et de composants cellulaires constitués d'adipocytes, de fibroblastes et de cellules immunitaires (Connor *et al.*, 2007).

La lactogenèse débuterait deux à trois jours avant la parturition. Cette phase est marquée par l'apparition de l'expression des ARNm de l'α-lactalbumine, une protéine du lactosérum impliquée dans la synthèse du lactose (Akers, 1985).

I.2.2. Régulation hormonale des différentes étapes de la biologie de la glande mammaire

Le contrôle de la mise en place des structures mammaires et de leur fonctionnement est assuré par l'action combinée de plusieurs hormones et facteurs locaux qui agissent simultanément ou de manière séquentielle et dans des rapports de concentration bien définis (Jammes & Djiane, 1988).

Ainsi, la croissance allométrique pré-pubertale est régulée par des estrogènes et l'hormone de croissance. La phase folliculaire des cycles sexuels implique les estrogènes alors que la phase lutéale nécessite en plus la progestérone (Figure I.9). La prolifération des canaux en début de gestation est due aux estrogènes, à la prolactine et à l'hormone de croissance. La formation des structures lobulo-alvéolaires, en gestation, implique les estrogènes, la progestérone et l'hormone de croissance. En fin de gestation, la prolifération cellulaire, est régulée par les œstrogènes, la progestérone, l'hormone de croissance et, dans une moindre mesure, l'hormone placentaire lactogène. A la parturition, la différenciation cellulaire et l'activation de la sécrétion sont induites par la prolactine, les estrogènes, l'hormone de croissance, les glucocorticoïdes et l'hormone placentaire lactogène. Le maintien de la lactation implique la prolactine, l'hormone de croissance, l'ocytocine, les glucocorticoïdes et l'insuline.

Les paragraphes suivant décrivent plus précisément ces hormones et leur(s) action(s) locale(s) sur les tissus constituant la glande mammaire.

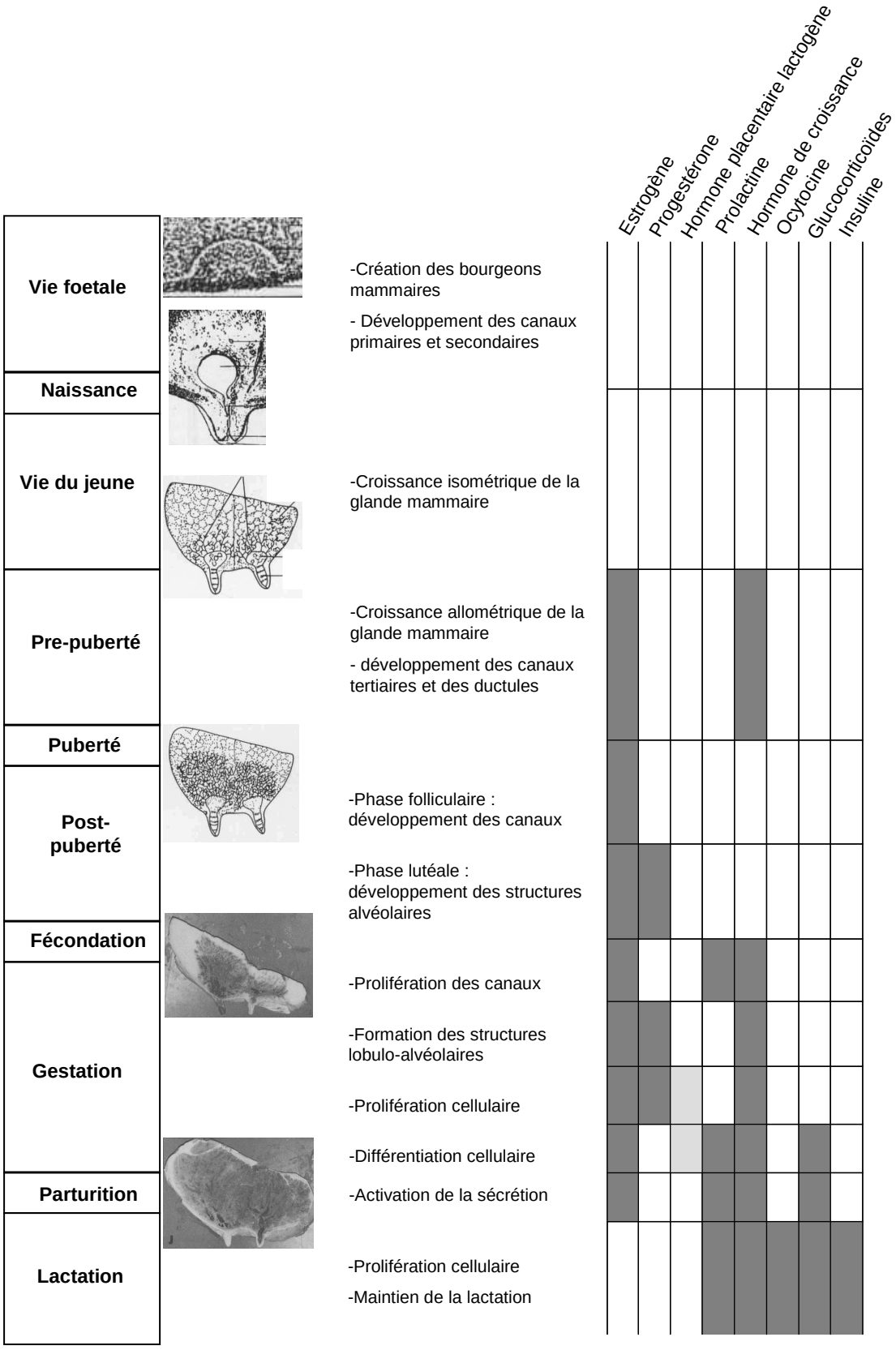


Figure I-9 : Représentation chronologique du développement et de la différenciation de la glande mammaire bovine au cours de la vie d'une femelle et régulations hormonales de ces étapes de développement (adapté de Jammes & Djiane, 1988; Neville *et al.*, 2002).

I.2.2.1. Les hormones de la reproduction

Les hormones de la reproduction interviennent dans la mise en place et le fonctionnement des organes de la reproduction tels que les organes génitaux et la glande mammaire.

➢ Les hormones stéroïdiennes

Les hormones stéroïdiennes sont d'origine ovarienne ou placentaire pendant la gestation. Elles sont essentielles à la mammogenèse (Jammes & Djiane, 1988).

Chez les ruminants, les estrogènes régulent la croissance allométrique pré-pubère de la glande mammaire et la croissance des canaux mammaires pendant la phase folliculaire des cycles sexuels, en stimulant la prolifération des cellules épithéliales mammaires, des cellules endothéliales et des fibroblastes (cf. revues de Connor *et al.*, 2007; Hovey *et al.*, 2002). Des injections d'estrogènes rétablissent le développement des canaux chez des génisses ovariectomisées. Chez la souris gestante, les estrogènes sont nécessaires à la ramification des canaux et au développement alvéolaire. Les estrogènes peuvent aussi induire la synthèse des récepteurs à la progestérone dans les cellules épithéliales mammaires (Hovey *et al.*, 2002). Juste après la parturition, les estrogènes agissent directement sur le tissu mammaire en stimulant la synthèse des caséines et en augmentant le nombre de récepteurs à la prolactine (Jammes & Djiane, 1988). Cependant, la lactation atteint son niveau maximal lorsque la concentration plasmatique d'œstrogène est basse (Erb, 1977). Les estrogènes seuls ne peuvent induire une croissance normale et complète des tissus sécréteurs mammaires (Hovey *et al.*, 2002).

L'évolution des concentrations en hormones stéroïdiennes plasmatiques au cours du cycle oestral chez les bovins est schématisée à la Figure I.10. La progestérone qui provient du corps jaune de l'ovaire, supprime l'effet mitogène des estrogènes (Hovey *et al.*, 2002). En phase lutéale, l'association estrogène (en faible concentration) et progestérone (en forte concentration) stimule la mise en place du système lobulo-alvéolaire (Jammes & Djiane, 1988). La progestérone est essentielle pour la différenciation lobulo-alvéolaire, la prolifération alvéolaire et l'activation de la synthèse des constituants du lait dès la mi-gestation (Connor *et al.*, 2007; Neville *et al.*, 2002). Pendant la gestation, la progestérone est sécrétée abondamment par l'unité foeto-placentaire chez la vache, la truie et la brebis mais pas chez la chèvre (Jammes & Djiane, 1988). La progestérone inhibe la lactogenèse en empêchant la décharge de prolactine au niveau hypophysaire et en diminuant la capacité de la

prolactine à induire la biosynthèse des protéines du lait au niveau de la glande mammaire (Jammes & Djiane, 1988). La chute de la concentration de progestérone en fin de gestation est le déclencheur de la sécrétion lactée (Neville *et al.*, 2002). La progestérone inhibe l'initiation de la lactation mais n'inhibe pas une lactation déjà établie (Akers, 1985).

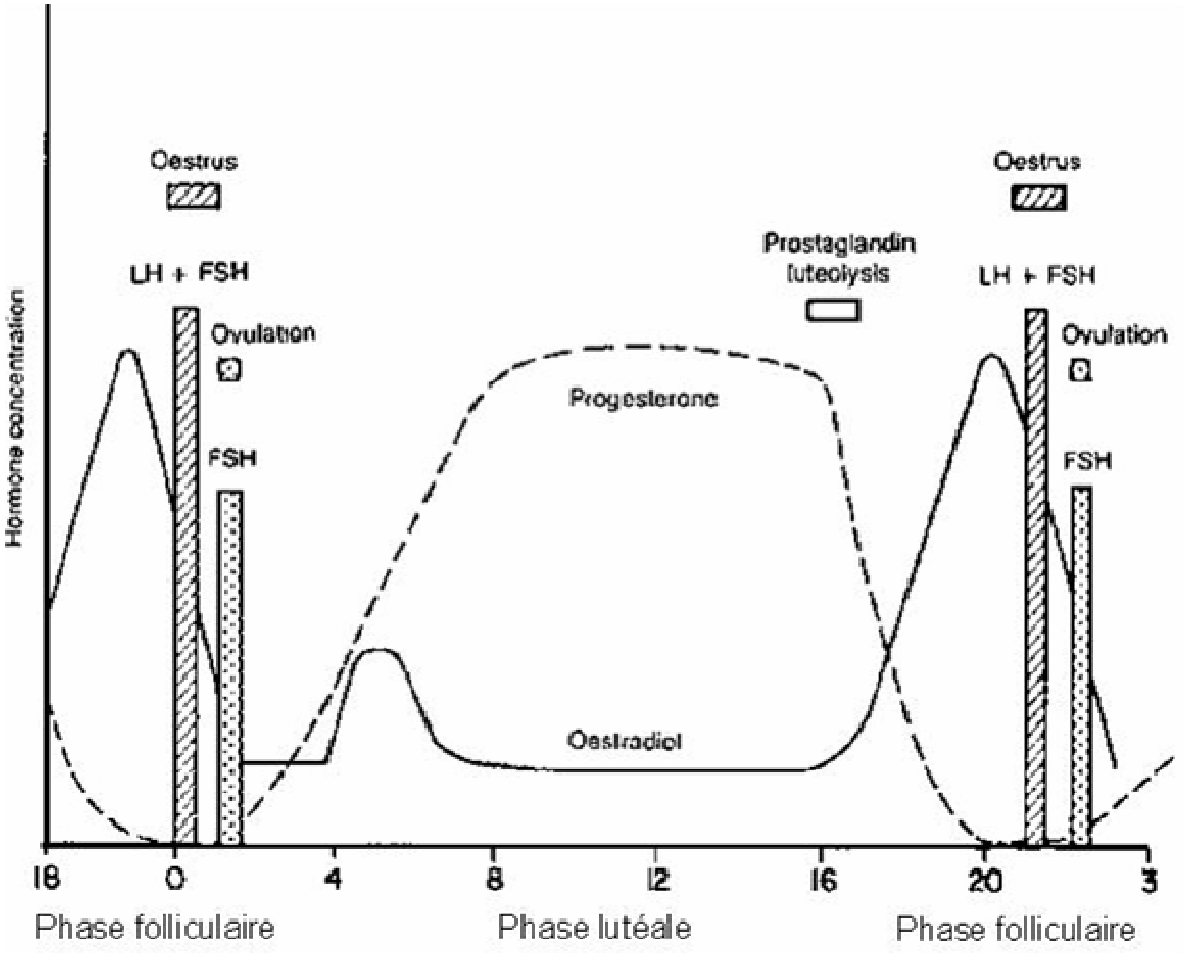


Figure I-10 : Représentation schématique des changements de concentration des hormones stéroïdiennes plasmatiques lors du cycle oestral bovin (adapté de Peters & Lamming, 1983)

Les sites de liaison spécifiques des hormones stéroïdiennes ont une localisation nucléaire (Jammes & Djiane, 1988) avec l'« Estrogen Receptor alpha », ERα et le « Progesterone receptor B », PR-B (Connor *et al.*, 2007). Les voies de régulation activées par les hormones stéroïdiennes sont relativement mal connues. Connor *et al*. (2007) proposent toutefois le modèle d'action suivant : l'estrogène et la progestérone se lieraient à des récepteurs situés dans l'épithélium (respectivement ERα et PR-B,), ce qui stimulerait la prolifération des cellules épithéliales proches via des facteurs dérivés de l'épithélium (« Epidermal Growth Factor », EGF et Wnt-4) et via des facteurs dérivés du stroma (« Fibroblast Growth Factor », FGF et « Insulin-like Growth Factor », IGF-I). La chute de la

progestérone à la parturition permettrait à la prolactine et à l'IGF-I de stimuler les sécrétions lactées via l'activation d'*Akt1* (« v-akt murine thymoma viral oncogene homolog 1 », une protéine kinase). La synthèse lipidique serait activée via la stimulation de *SREBP1* (également *SREBF1*, « sterol regulatory element binding transcription factor 1 ») et d'autres gènes spécifiant des protéines impliquées dans la biosynthèse des lipides.

L'hormone placentaire lactogène (PL), secrétée par le placenta, est aussi appelée la sommatomammotropine chorionique ou mammotropine chorionique. La quantité de PL augmente nettement dans le sang dès la moitié de la gestation chez les chèvres et les brebis, mais pas chez les vaches. Cette hormone n'est pas indispensable au développement normal de la glande mammaire ou à la lactation. Elle interviendrait dans le développement de la glande mammaire des chèvres et des brebis en permettant la différenciation et la prolifération des cellules épithéliales des alvéoles ce qui entraînerait une augmentation de la production laitière (Akers, 1985; Hennighausen & Robinson, 2001).

➢ Les hormones hypophysaires

Les hormones hypophysaires ont un rôle d'amplificateur de l'action des hormones stéroïdiennes pendant la mammogenèse (Jammes & Djiane, 1988). La prolactine est secrétée par l'hypophyse antérieure. Pendant la puberté, la croissance de la glande mammaire et la concentration plasmatique de prolactine sont corrélées positivement (Hovey *et al.*, 2002), ce qui indique que cette hormone intervient dans la croissance de la glande mammaire. Pendant la gestation, cette hormone est indispensable à la prolifération et la différenciation fonctionnelle des structures lobulo-alvélaires (Hennighausen & Robinson, 1998; Neville *et al.*, 2002). Chez la brebis, l'interaction entre prolactine et estrogène joue un rôle fondamental pour le développement alvéolaire, alors que seule la prolactine semble nécessaire chez la génisse (Hovey *et al.*, 2002). Pendant la gestation, tant que la concentration en progestérone est élevée, la prolactine ne peut initier la synthèse d'α-lactalbumine et de lactose, même si sa concentration dans le plasma et les sécrétions mammaires augmente avant que la concentration de progestérone ne chute (Erb, 1977). A la parturition, la prolactine est indispensable pour l'activation et le maintien de la sécrétion lactée (Neville *et al.*, 2002). Elle a une action directe sur l'épithélium alvéolaire (Akers, 1985) via le récepteur de la prolactine, qui appartient à la superfamille des récepteurs de cytokines (Bole-Fyesot *et al.*, 1998). La liaison entre la prolactine et son récepteur entraîne la dimérisation du récepteur et la phosphorylation par la Janus Kinase 2 (Jak2) d'un résidu tyrosine spécifique associé au récepteur. Par la suite, les facteurs de transcription Stat5a et Stat5b sont mobilisés par leurs

domaines SH_2 et sont phosphorylés par Jak2. Une fois phosphorylés, Stat5a et Stat5b forment des homo et hétérodimers et sont dirigés vers le noyau où ils activent la prolifération et la différenciation cellulaire. Pendant la gestation, Stat5a est limitant pour la détermination, la prolifération et la différenciation des alvéoles mammaires (Hennighausen & Robinson, 2001).

L'hormone de croissance (ou « growth hormone », GH) est sécrétée par l'hypophyse antérieure. La liaison entre la GH et son récepteur localisé sur les cellules du stroma et les hépatocytes, entraîne la sécrétion d'IGF-I, qui agit sur les cellules épithéliales mammaires en activant la mitose, comme décrit chez la brebis pendant la croissance allométrique pré-pubère (Hovey *et al.*, 2002), ou en régulant leur survie (Neville *et al.*, 2002). La GH participe également à la lactation en favorisant la mobilisation des graisses par inhibition de la lipogenèse induite par l'insuline et par stimulation de la lipolyse dans les adipocytes (Peel & Bauman, 1987). Elle intervient aussi dans la redistribution des nutriments en faveur de la glande mammaire et augmente le flux sanguin mammaire (Johnson & Hart, 1986) cité par (Jammes & Djiane, 1988).

L'ocytocine est une hormone secrétée par l'hypophyse postérieure. Elle intervient dans l'éjection du lait en provoquant la contraction des cellules myoépithéliales entourant les alvéoles par interaction avec des récepteurs spécifiques situés sur ces cellules (Neville *et al.*, 2002). Elle intervient aussi dans le développement de la glande mammaire post-partum en activant la prolifération des alvéoles (Hennighausen & Robinson, 1998).

I.2.2.2. Les hormones métaboliques

Les hormones métaboliques sont responsables de la coordination de la réponse corporelle aux changements métaboliques et au stress (Neville *et al.*, 2002).

Les glucocorticoïdes sont des hormones sécrétées par la glande corticosurrénale. Le cortisol est le glucocorticoïde prédominant chez la vache. Sa fonction majeure dans la glande mammaire est d'entraîner la différenciation du système lobulo-alvéolaire au moment de la délivrance. Les glucocorticoïdes et la prolactine agissent en synergie pour l'activation de la lactation (Tucker, 2000). La quantité de glucocorticoïdes augmente à la parturition, probablement en lien au stress de la mise-bas (Erb, 1977) et est souvent élevée pendant la lactation (Neville *et al.*, 2002).

L'insuline est une hormone sécrétée par les cellules β du pancréas endocrine. Elle interviendrait au niveau de la glande mammaire dans la lipogenèse et la synthèse du lactose, en régulant l'apport de nutriments à la glande mammaire pendant la lactation. Son taux sanguin est relativement bas pendant la lactation, ce qui conduirait à limiter le stockage de

nutriment dans les réserves corporelles, afin de les rendre disponibles pour la synthèse lactée (Neville *et al.*, 2002).

I.2.2.3. Variations des concentrations hormonales ante et post partum

La concentration de progestérone dans le sang diminue 2-3 jours avant le vêlage alors que les concentrations en prolactine et GH augmentent 1 à 2 jours avant le vêlage et restent élevées 1 à 2 jours après le vêlage (Akers, 1985; Erb, 1977). Pendant le mois qui suit la mise-bas, la concentration sanguine en GH diminue et la concentration d'insuline augmente. La concentration en glucocorticoïdes sériques augmente significativement pendant le vêlage, probablement en lien avec le stress de la parturition et diminue dans les 12h qui suivent. Dans le plasma, le sérum sanguin, l'urine ou les sécrétions mammaires, la concentration en estrogènes augmente rapidement pendant la première semaine de lactation. Dans le sang, elle est encore maximale un jour avant le vêlage et le reste un jour après le vêlage dans l'urine ou le lait (Erb, 1977).

I.2.2.4. Facteurs locaux intervenant dans la mammogenèse et la lactogenèse

L'IGF-I est un facteur de croissance sécrété par les cellules du foie ou du stroma après activation de celles-ci par la GH. Dans la glande mammaire ovine, la plus grande expression du gène spécifiant IGF-I a été observée, au niveau des transcrits, lors de la croissance allométrique pré-pubère et lors du développement alvéolaire en fin de gestation (Hovey *et al.*, 1999). Parmi les facteurs de croissance, IGF-I et IGF-II régulent la mammogenèse et la galactopoïèse chez la vache et la brebis (Akers, 1990). Chez la génisse entre la naissance et la puberté, l'expression d'IGF-I est plus élevée dans le tissu adipeux que dans le parenchyme mammaire et cette expression coïncide avec une prolifération des cellules épithéliales (Connor *et al.*, 2007).

Les autres facteurs sécrétés soit par le stroma soit par l'épithélium et intervenant dans le développement de la glande mammaire ou la formation des alvéoles ont été décrits chez la souris et sont le facteur de différenciation des cellules osseuses (RANKL), l'inhibine βB et les familles des TGFβ (Hennighausen & Robinson, 2001), la cycline D1, A-myb, C/EBPβ (Hennighausen & Robinson, 1998). La cycline D1 participerait au développement alvéolaire en activant le cycle cellulaire. En effet, les souris pour lesquelles le gène de la cycline D1 a été invalidé (« knock out », KO) ne forment pas d'alvéoles fonctionnelles pendant la gestation

(Hennighausen & Robinson, 1998). De plus, ces mêmes auteurs rapportent que les souris dont les gènes A-myb et C/EBPβ ont été inactivés, présentent un développement et une différenciation alvéolaire sévèrement réduits pendant la gestation.

Par ailleurs, il a été montré que la glande mammaire pouvait aussi produire des hormones telles que la GH, la prolactine, la leptine et la PTHrP (« Parathyroid hormone related protein »). Ces quatre hormones sont présentes dans le lait de souris (Neville *et al.*, 2002). La signalisation entre l'épithélium et le mésenchyme via la PTHrP et son récepteur est un noeud important dans l'élongation de la première ébauche de glande mammaire pendant la vie fœtale (Hennighausen & Robinson, 2001).

I.3. Composition et biosynthèse des constituants majeurs du lait

Le lait est le premier aliment d'un jeune mammifère. C'est un aliment complet dont la composition évolue et est adaptée à la survie, la croissance et au développement du nouveau-né. Le lait, est un sérum comportant une émulsion de matière grasse, une suspension de matière protéique caséeuse, du lactose, des sels et minéraux, des protéines solubles et des traces d'éléments divers. Les laits sécrétés par différentes espèces de mammifères présentent des caractéristiques communes et contiennent les mêmes familles de constituants. Cependant, les proportions respectives de ces différents constituants varient largement d'une espèce à l'autre en lien avec leur mode de vie et de reproduction (Tableau I.1). La teneur en matière grasse varie de 1,4% (ânesse) à plus de 50 % (phoque) ; la teneur en glucides, principalement le lactose, varie de traces (phoque) à 8% (ânesse) ; la teneur en protéines varie de 1% (femme) à près de 15% (oursonne) et la teneur en minéraux varie de 0,2% chez la femme à 1,8% chez l'oursonne (Jenness, 1986).

Tableau I-1 : Teneurs moyennes des principaux constituants du lait de différentes espèces[1]

Espèces		Eau (%)	Matière grasse (%)	Matière protéique (%) (Caséines + Protéines du lactosérum)	Lactose (%)	Minéraux (%)	Energie (kcal/100g)
Anesse	*Equus asinus*	88,3	1,4	2 (1 + 1)	7,4	0,5	44
Brebis	*Ovis aries*	82	7,2	4,6 (3,9 + 0,7)	4,8	0,9	102
Bufflonne	*Bubalus bubalis*	82,8	7,4	3,8 (3,2 + 0,6)	4,8	0,8	101
Chamelle	*Camelus dromedarius*	86,5	4	3,6 (2,7 + 0,9)	5	0,8	70
Chevre	*Capra hircus*	86,7	4,5	3,2 (2,6 + 0,6)	4,3	0,8	70
Femme	*Homo sapiens*	87,1	4,5	0,9 (0,4 + 0,5)	7,1	0,2	72
Jument	*Equus caballus*	88,8	1,9	2,5 (1,3 + 1,2)	6,2	0,5	52
Ourse Noire	*Ursus americanus*	55,5	24,5	14,5 (8,8 + 5,7)	0,4	1,8	280
Phoque	*Callorbinus ursinus*	34,6	53,3	8,9 (4,6 + 4,3)	0,1	0,5	516
Rate	*Rattus norvegicus*	72,4	9,3	8,2 (7,2 + 1)	3,7	1,4	131
Renne	*Rangifer tarandus*	66,7	18	10,1 (8,6 + 1,5)	2,8	1,5	214
Vache	*Bos taurus*	87,3	3,9	3,2 (2,6 + 0,6)	4,6	0,7	66

[1] d'après Jenness, 1986

I.3.1. Les glucides

Le principal glucide du lait de la plupart des mammifères est le lactose ou 4-O-β-D-galactopyrannosyl-D-glucopyrannose. C'est un disaccharide composé de β-D-galactose et de β-D-glucose. Les autres glucides du lait sont soit des glucides modifiés, c'est-à-dire liés à des phosphates, des lipides ou des protéines, soit des oligoholosides libres tels que le glucose et le galactose (Holsinger, 1997; Pougheon & Goursaud, 2001). La teneur en glucides, variable au cours de la lactation, est différente selon l'espèce : par exemple, le lait humain contient beaucoup plus de glucides autres que le lactose que le lait de vache, respectivement 1,3 et 0,3 % (Pougheon & Goursaud, 2001).

La réaction de synthèse du lactose (Figure I.11) nécessite une molécule de glucose et une molécule de galactose liées entre elles par une liaison osidique β(1→4).

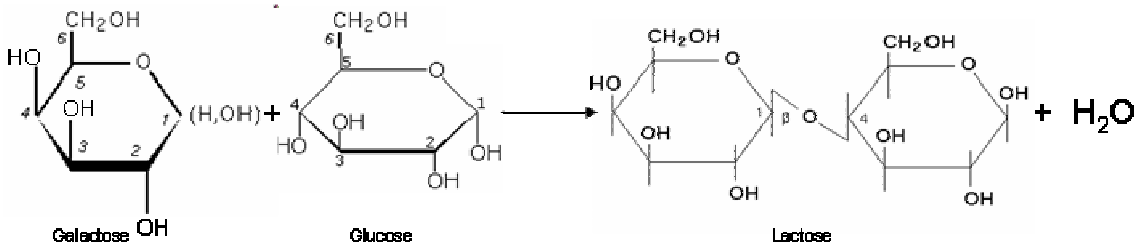


Figure I-11 : Réaction de synthèse du lactose

Le glucose est capté de la circulation sanguine par la CEM via le transporteur GLUT1 ancré dans la membrane basale (1A sur la Figure I.12). Chez les ruminants, le glucose sanguin est synthétisé dans le foie selon un processus appelé néoglucogenèse (Figure I.13). Les substrats de la néoglucogenèse sont le propionate et les acides aminés glucogéniques (tous les

acides aminés sauf la leucine). Le propionate provient de la fermentation des glucides par les micro-organismes du rumen (Pougheon & Goursaud, 2001).

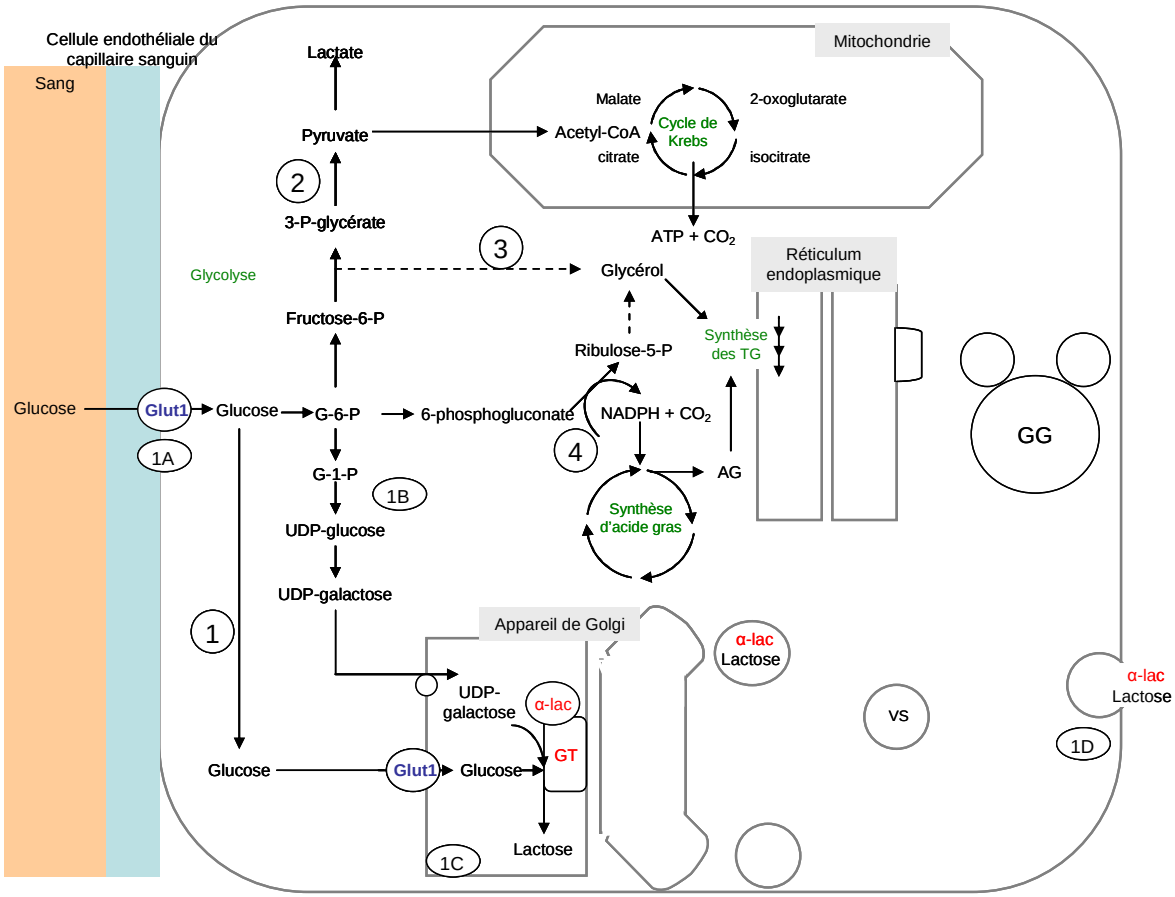


Figure I-12 : Utilisation du glucose par la cellule épithéliale mammaire et voie de synthèse du lactose (d'après Bernard *et al.*, 2008; Guinard-Flament *et al.*, 2006; Kuhn *et al.*, 1980)

1 : voie de synthèse du lactose ; 2 : glycolyse ; 3 : synthèse du glycérol ; 4 : voie de synthèse pentose-phosphate. 1A : captation du glucose par la cellule épithéliale mammaire ; 1B : voie de synthèse de l'UDP-galactose ; 1C : réaction de synthèse du lactose ; 1D : sécrétion du lactose. G-6-P : Glucose-6-phosphate ; G-1-P : Glucose-1-phosphate ; α-lac : α-lactalbumine ; GT : UDP-Galactosyltransferase ; Fructose-6-P : Fructose-6-phosphate ; 3-P-glycérate : 3-phosphate-glycérate ; Ribulose-5-P : Ribulose-5-phosphate ; AG : acides gras ; TG : triglycérides ; VS : vésicule de sécrétion ; GG : globule gras.

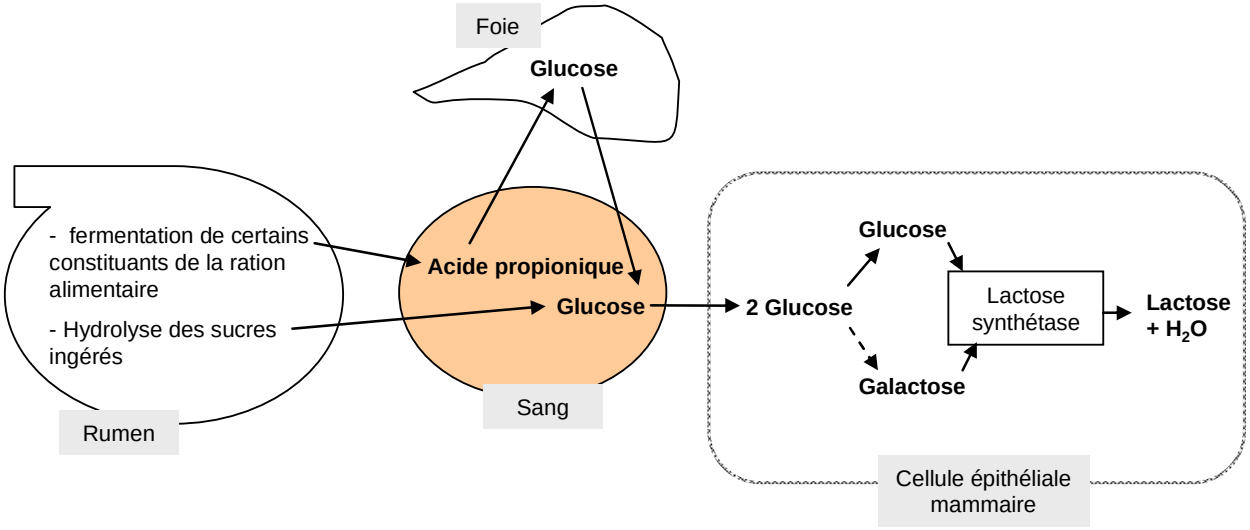


Figure I-13 : Origine du glucose capté par la cellule épithéliale mammaire (d'après Pougheon & Goursaud, 2001)

Dans la CEM, la molécule de galactose est synthétisée à partir d'une molécule de glucose (voie 1B sur la Figure I.12) en passant par les étapes de glucose-6-phosphate puis glucose-1-phosphate et enfin UDP-glucose, faisant intervenir respectivement les enzymes hexokinase, phosphoglucomutase et UDP-glucose pyrophosphorylase (Ebner, 1971).

Au niveau de la CEM, les deux tiers du glucose capté servent à la synthèse du lactose qui a lieu dans l'appareil de Golgi (voie 1C sur la Figure I.12). Le glucose est transporté dans ce compartiment sans consommation d'énergie via le transporteur de glucose (Glut1) situé, pendant la lactation, dans la membrane de l'appareil de Golgi (Nemeth *et al.*, 2000). La molécule d'UDP-galactose est transportée de façon active (consommation d'énergie) dans la lumière de l'appareil de Golgi via les transporteurs UDP-GalT1 ou UDP-GalT2.

La synthèse de lactose, réaction non réversible, est catalysée par le complexe enzymatique lactose synthétase, composé de la β1,4-UDP galactosyltransferase (GT) et de l'α-lactalbumine (α-lac). La GT bovine est une glycoprotéine de 43 kDa ancrée dans la membrane de l'appareil de Golgi. En l'absence d'α-lac, la GT intervient dans la biosynthèse des glycoprotéines en catalysant l'ajout d'un galactose aux motifs oligosaccharidiques porteurs d'un résidu N-glucosamine terminal. L'α-lac est une protéine du lait, non catalytique, nécessaire à la synthèse de lactose. Elle altère la spécificité de substrat de la GT vers le glucose au lieu du résidu N-acetylglucosamyl, et donc oriente son activité enzymatique vers la synthèse de lactose (Holsinger, 1997). Le lactose formé ne diffuse pas hors de l'appareil de Golgi. De l'eau entre dans les vésicules de sécrétion de façon à équilibrer la pression

osmotique. L'α-lac et le lactose sont excrétés dans le lait par exocytose des vésicules de sécrétion, conjointement à la sécrétion des protéines du lait (voie 1D sur la Figure I.12).

Nos connaissances sur la régulation de la synthèse du lactose, sont assez restreintes même si il y a déjà quarante ans, Palmiter concluait que la prolactine est la « gâchette » qui déclenche la synthèse du lactose dans le tissu mammaire murin, préalablement stimulé par l'insuline et l'hydrocortisone (Palmiter, 1969). Toutefois, une grande teneur en lactose dans l'appareil de Golgi n'inhibe pas sa biosynthèse. Le taux de synthèse du lactose dépendrait de la quantité et du ratio entre les deux sous unités de la lactose synthétase (α-lac et GT) dont la synthèse serait régulée par les hormones lactogènes (Holsinger, 1997). La diminution de la fréquence de traite et la restriction alimentaire provoque une inhibition de la synthèse de lactose (Guinard-Flament *et al.*, 2006).

Chez la vache laitière, le glucose non utilisé pour la synthèse de lactose servirait à la production : (I) d'énergie, notamment sous forme d'ATP (voie 2 sur la Figure I.12) ; (II) de glycérol, utilisé ensuite dans la synthèse des triglycérides (voie 3 sur la Figure I.12) ; (III) de ribose, employé pour la synthèse d'ARN et d'ADN ; (IV) de NADPH, généré par la voie de synthèse des pentoses phosphate et exploité dans la synthèse des acides gras (voie 4 sur la Figure I.12). Pour plus de détails, se reporter à l'excellente revue (Guinard-Flament *et al.*, 2006).

I.3.2. Les lactoprotéines

La teneur en protéines du lait varie suivant les espèces (cf. Figure I.14). Ainsi, le lait humain contient la plus faible teneur en protéines (10 g/L), alors que le lait de lapine est le plus riche en protéines (200 g/L). Les laits de vaches et de chèvres sont relativement pauvres en protéines (respectivement 33 et 30 g/L).

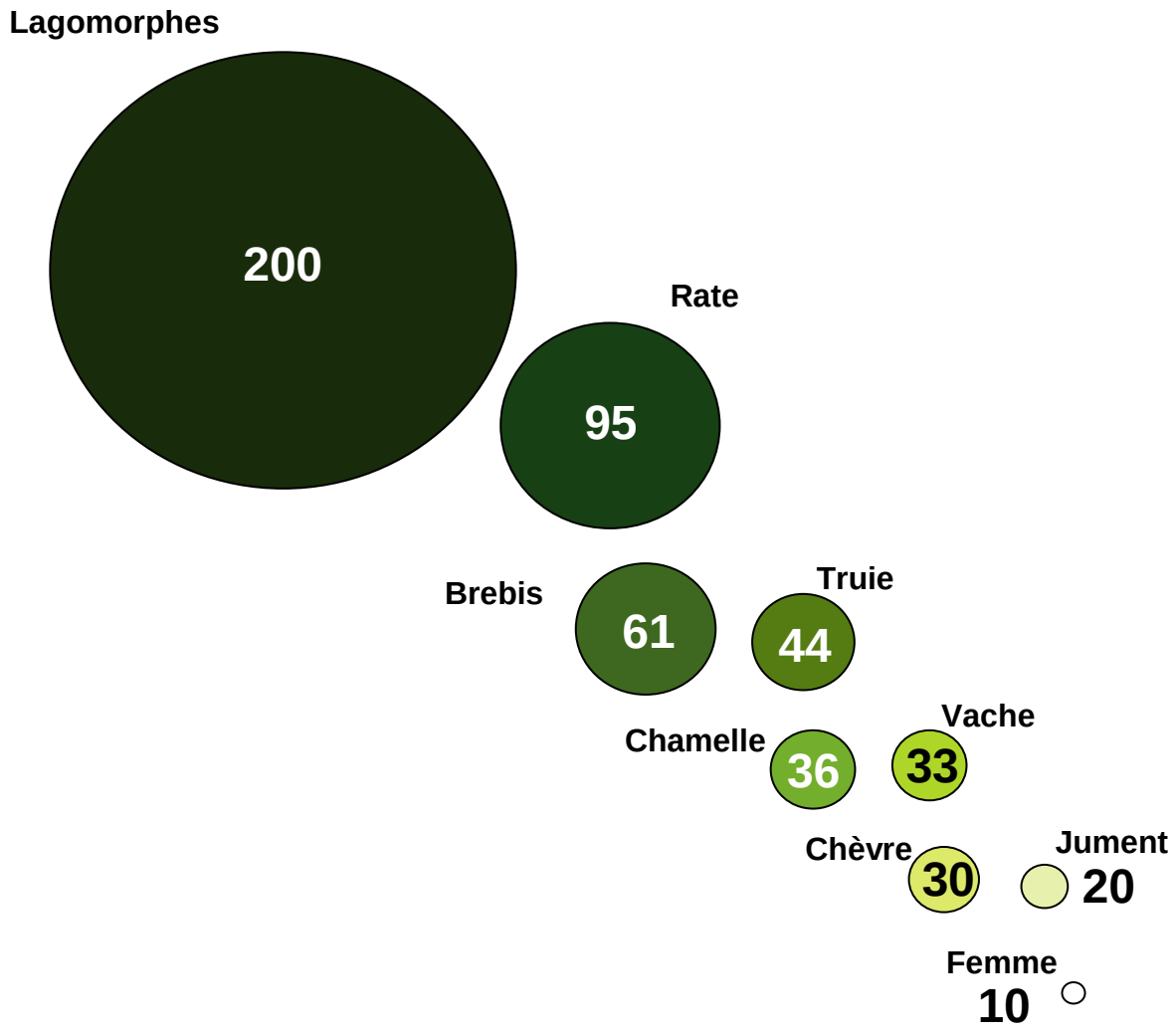


Figure I-14 : Teneur en protéines (g/l) du lait de neuf espèces (Martin *et al.*, 2003)

I.3.2.1. Composition protéique des laits

Le lait contient trois catégories de protéines : les protéines colloïdales formant les micelles de caséines, les protéines solubles du lactosérum et les protéines de la membrane du globule gras ou MFGM (cf. Tableau I.2).

Tableau I-2 : Concentration et caractéristiques des protéines du lait de vache[1]

Protéines	Concentration (g/L)	Point isolélectrique	Poids moléculaire
Protéines de la micelle de caséine			
Caséine α_{s1}	10	4,1	23 000
Caséine α_{s2}	2,6	4,1	23 000
Caséine β	9,3	4,5	24 000
Caséine κ	3,3	4,1	19 000
Caséine γ	0,8	5,8-6,0	
Protéines solubles du lactosérum			
Lactoglobuline β	3,2	5,3	18 000
α-lactalbumine	1,2	5,1	14 437
Serum albumine	0,4	4,7	68 000
Immunoglobuline	0,8		
Lactoferrine	0,1		87 000
Transferrine	0,1		
Proteines de la membrane du globule gras	0,4		
Autres	0,8		
Lait	33		

[1]adapté de Tremblay *et al.*, 2003

Les caséines sont des protéines synthétisées dans la cellule épithéliale mammaire, présentes dans le lait sous forme d'agrégats appelés micelles. Ces protéines coagulent à pH = 4,6. Ce sont des phosphoprotéines de poids moléculaire d'environ 20 kDa (cf. Figure I.15). Elles sont au nombre de quatre : les caséines α_{s1}, α_{s2}, β et κ. Les caséines α_{s1}, α_{s2} et β sont hydrophobes. La caséine β a la propriété de polymériser spontanément en solution. La caséine κ est soluble en présence d'ions calciques et c'est la seule caséine glycosylée. En solution, les caséines s'auto-assemblent pour former, en présence de calcium, des hétéro-polyméres que la caséine κ stabilise. L'ensemble de la structure constitue la micelle de caséines. Ces micelles sont donc constituées des quatre caséines auxquelles est lié de façon covalente du phosphore, sous forme de phospho-sérines (accessoirement phospho-thréonines), un complexe phosphocitrate d'ions calciques et magnésium amorphes et du calcium ionique lié aux caséines par des liaisons électrostatiques. Pour certains auteurs, les micelles de caséines seraient composées de sous-unités sphériques, les submicelles, de poids moléculaire de 61 kDa et de diamètre 15 à 200 nm (De Kruif & Holt, 2003). Ces sous-unités sont dépourvues de minéraux et sont cimentées entre elles par du phosphate et du citrate colloïdal avec un rapport (Ca^{2+}, Mg^{2+})/phosphate de 1,58. Cette vision de la structure micellaire est controversée et l'existence de submicelles est aujourd'hui contestée (Holt, 2004; Marchin *et al.*, 2007). La taille moyenne des micelles de caséines décrit des variations saisonnières et

dépend du stade et du numéro de lactation et de l'animal. Elle varie entre 216 et 428 nm dans le lait de vache (De Kruif & Holt, 2003).

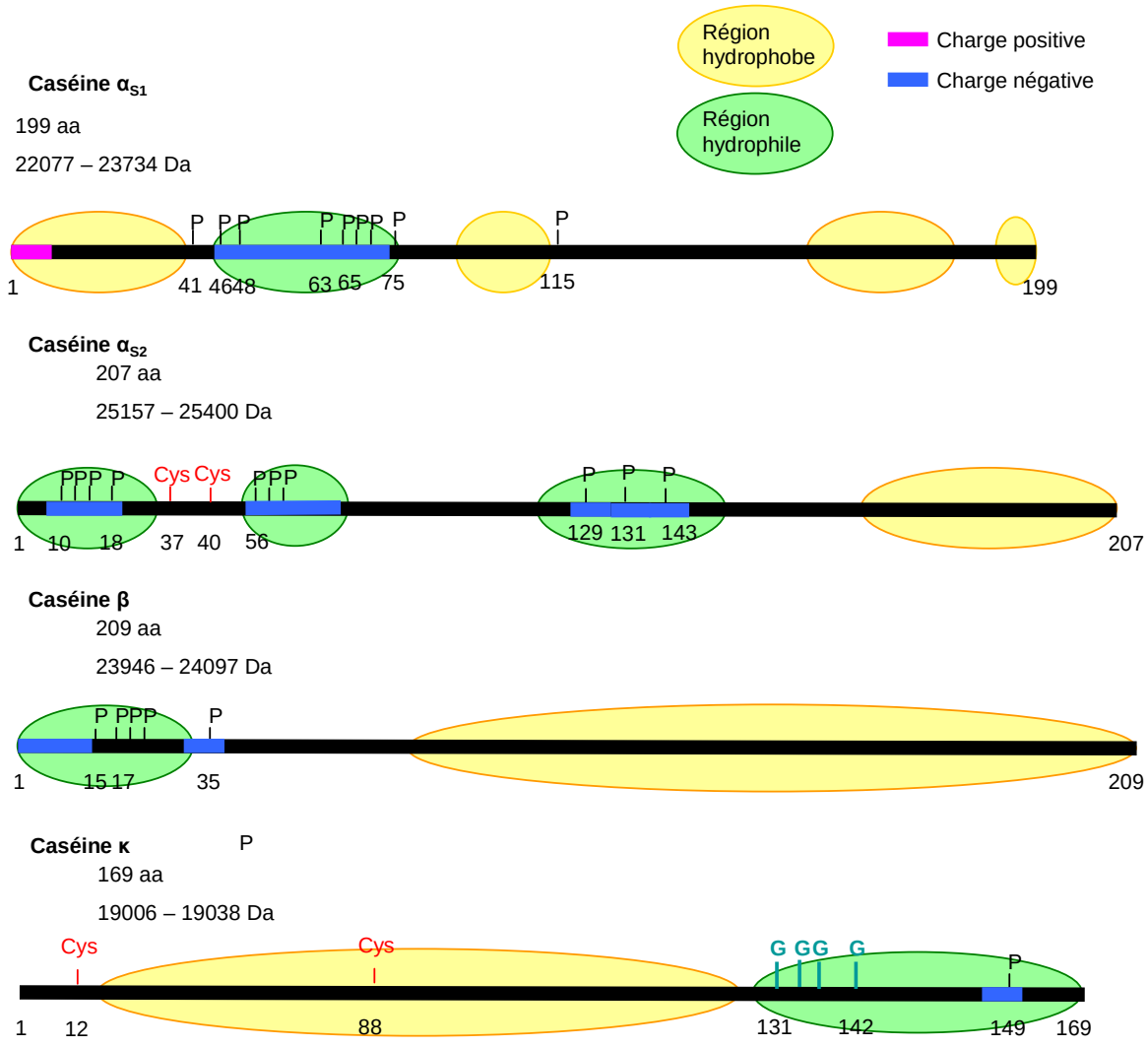


Figure I-15 : Quelques caractéristiques des caséines bovines (d'après Neveu, 2004; Swaisgood, 2003) aa : acides aminés ; Da : masse moléculaire en Daltons ; Cys : résidus cystéyles ; modifications post-traductionnelle (phosphorylations, P; Glycosylation, G) ; β : bêta, α_{s1} : alpha S1, α_{s2} : alpha S2, κ : kappa.

Les protéines du lactosérum possèdent une excellente valeur nutritionnelle, riches en acides aminés soufrés, en lysine, en tryptophane et sensibles à la dénaturation thermique. Les protéines majeures du lactosérum sont soit synthétisées dans la CEM (α-lactalbumine et β-lactoglobuline) ou proviennent directement du sang (sérum-albumine et immunoglobulines). Certaines protéines du lactosérum telles que la lactopéroxydase et la xanthine oxidase possèdent une activité enzymatique. D'autres sont des capteurs d'ion $Fe^{2+/3+}$ telles la lactoferrine et la transferrine.

Les protéines de la MFGM constitue une classe à part entière de la fraction protéique du lait, souvent hautement hydrophobes et peu solubles (O'Donnell *et al.*, 2004). Les protéines majeures de la MFGM sont la mucine 1 (MUC1), la xanthine oxidase (XDH), PASIII (une glycoprotéine), le CD36, la butyrophiline (BTN), l'adipophiline (ADRP ou ADFP) et la « Fatty Acid Binding Protein » (FABP) (pour revue Mather, 2000). MUC1 est une protéine membranaire de haut poids moléculaire, très glycosylée et incorporée dans la MFGM lors du bourgeonnement et expulsion du globule gras par la surface cellulaire. Il est suggéré que MUC1 a un rôle immunoprotectif. L'enzyme XDH jouerait un rôle structural et fonctionnel dans la formation de la MFGM en se liant au domaine cytoplasmique de la BTN, formant ainsi un complexe protéique qui permettrait l'interaction entre la gouttelette lipidique et la membrane plasmique apicale. Le CD36 lient les acides gras à longue chaîne et agit comme un transporteur d'acides gras. La BTN est la protéine la plus abondante de la MFGM bovine, elle participerait à la sécrétion du globule gras par la CEM. L'ADRP permettrait la captation des acides gras à longue chaîne et l'accrétion et le transport intracellulaire des gouttelettes lipidiques. Cette protéine est liée à la membrane de la gouttelette lipidique et participerait à la sécrétion du globule gras par la CEM en se liant au complexe XDH/BTN situé dans la membrane plasmique apicale (Mather & Keenan, 1998b). Dans le tissu mammaire, la FABP interviendrait dans le transport intracellulaire des acides gras, le contrôle du métabolisme lipidique et l'accrétion des gouttelettes lipidiques dans le cytoplasme.

Les teneurs relatives des composants du lait dépendent, pour partie, des variants génétiques aux protéines majeures. Dans l'espèce bovine, 8, 4, 13, 12, 11 et 3 variants génétiques ont été identifiés pour, respectivement, les caséines α_{s1}, α_{s2}, β et κ, β-lactoglobuline et l'α-lactalbumine (Tableau I.3, Farrell *et al.*, 2004). Dans l'espèce caprine, 16, 7, 9 et 16 variants génétiques ont été rapportés pour respectivement, les caséines α_{s1}, α_{s2}, β et κ (Chessa *et al.*, 2008a; Chessa *et al.*, 2008b; Chiatti *et al.*, 2007; Neveu *et al.*, 2002). Pour cette espèce, il a été détecté 15 mutations ponctuelles au niveau du gène spécifiant la β-lactoglobuline (Ballester *et al.*, 2005) et deux variants de α-lactalbumine ont été identifié (Cosenza *et al.*, 2003). Pour les ovins, 5, 4, 2, 2, 3 et 2 variants génétiques ont été identifiés pour, respectivement, les caséines α_{s1}, α_{s2}, β et κ, la β-lactoglobuline et l'α-lactalbumine (Ceriotti *et al.*, 2004; Chianese *et al.*, 1996; Ng-Kwai-Hang & Grosclaude, 2003; Recio *et al.*, 1997).

Dans l'espèce bovine, les individus homozygotes BB pour la caséine κ produisent du lait caractérisé par un taux protéique et des quantités sécrétées de matière protéique et caséine κ plus élevés (Ng-Kwai-Hang *et al.*, 1984).

Les 16 variants détectés au niveau du locus de la caséine α_{s1} (*CSN1S1*) chez la chèvre (Chessa *et al.*, 2008a; Neveu *et al.*, 2002) sont associés à 4 niveaux de synthèse différents de la protéine (Grosclaude *et al.*, 1994) : un taux « fort » (3,6 g/l/allèle ; allèle A, B1, B2, B3, B4, B', C, H, L et M), un taux « moyen ou intermédiaire » (1,6 g/l/allèle ; allèles I et E), un taux « faible » (0,6 g/l/allèle ; allèles F et G), et jusqu'à une absence totale de synthèse de caséine α_{s1} (0 g/l/allèle) pour les allèles O1, O2, O3 et N. Pour ces derniers variants génétiques, il a été reporté, dans la littérature, un dysfonctionnement de la CEM (Chanat *et al.*, 1999).

Tableau I-3 : Variants génétiques aux protéines majeures des principaux ruminants

	caséine α_{s1} *(CSN1S1)*	caséine α_{s2} *(CSN1S2)*	caséine β *(CSN2)*	caséine κ *(CSN3)*	β-lactoglobuline *(BLG)*	α-lactalbumine *(LALBA)*
Bovins	8 variants[i] A à H	4 variants[i] A à D	13 variants[i] A1, A2, A3, A4, B à G, H1, H2 et I	12 variants[i] A à E, F1, F2, G1, G2, H à J	11 variants[i] A à J et W	3 variants[i] A à C
Caprins	18 variants[d,j] A, B1, B2, B3, B4, B', C, E à I, L à N, O1, O2, O3	7 variants[d] A à F et O	9 variants[e] A, A1, C, E, O, O', B, D	16 variants[g] A, B, B', B'', C, C', D à M	30 variants[a]	2 variants[h]
Ovins	5 variants[f] A à E	4 variants[l] A à C + ?	2 variants[c]	2 variants[c]	3 variants[k] A à C	2 variants[k] A et B

a : Ballester *et al.*, 2005 ; b : Caroli *et al.*, 2007 ; c : Ceriotti *et al.*, 2004 ; d : Chessa *et al.*, 2008a ; e : Chessa *et al.*, 2008b ; f : Chianese *et al.*, 1996 ; g : Chiatti *et al.*, 2007 ; h : Cosenza *et al.*, 2003 ; i : Farrell *et al.*, 2004 ; j : Neveu *et al.*, 2002 ; k : Ng-Kwai-Hang & Grosclaude, 2003 ; l : Recio *et al.*, 1997.

I.3.2.2. Origine des acides aminés servant à la synthèse des protéines du lait

Les acides aminés (AA) servant à la synthèse des protéines du lait sont soit captés du sang par la CEM soit synthétisés *de novo* dans la CEM. Les AA présents dans le sang viennent de la fermentation dans le rumen des fibres apportées par l'alimentation. Certains acides aminés essentiels (AAE) : la Lysine, la Méthionine, la Phénylalanine, la Tyrosine et le Tryptophane, sont captés en quantité stoechiométrique (1 pour 1) alors que d'autres (AA ramifiés et Arginine) sont souvent captés en quantité supérieure à la teneur nécessaire pour la synthèse des lactoprotéines. Les AA non essentiels (AANE) sont souvent captés en quantité inférieure à la teneur nécessaire pour la production des protéines du lait. Les AAE captés en excès sont soit utilisés comme substrat pour la synthèse d'AANE (Mepham *et al.*, 1992), soit

catabolisés pour la production d'énergie indispensable au fonctionnement de la cellule. Le passage des AA dans la CEM est facilité par la présence de transporteurs membranaires de nature protéique. Ces transporteurs ont une affinité pour plusieurs AA. Ils sont régulés par les hormones lactogènes ou les ions.

I.3.2.3. Mécanisme de synthèse des protéines du lait

Le mécanisme de synthèse des protéines du lait est semblable aux mécanismes de synthèse des protéines secrétées dans une cellule sécrétrice eucaryote. Les étapes de ce processus sont les suivantes et peuvent être successives ou simultanées : un gène est transcrit en ARN pré-messager (ARNpm) dans le noyau de la cellule, l'ARNpm mature pour devenir un ARN messager (ARNm, epissage) qui est traduit en protéine au niveau du réticulum endoplasmique granuleux (REG). Lors de cette traduction la chaîne polypeptidique nouvellement formée passe dans la lumière du REG et de l'appareil de Golgi où elle peut subir des modifications co- ou post-traductionnelles (glycosylation, phosphorylation…).
La synthèse réalisée dans cette partie est inspirée des revues sur la biosynthèse des protéines publiées par Mepham *et al.*, 1992; Mepham *et al.*, 1982; Vilotte *et al.*, 2003.

➢ Du gène à l'ARNm

Chez les Eucaryotes, la transcription des gènes spécifiant des protéines est contrôlée de façon précise et fait intervenir une enzyme : l'ARN polymérase II (Pol II), et de nombreux facteurs de transcription. Cette enzyme est constituée de 8 à 10 sous-unités et est capable de polymériser les ribonucléotides en ARNpm par lecture du brin d'ADN dans le sens 5'→ 3'. Les systèmes de régulation de la transcription sont nombreux et peuvent se situer à la fois au niveau de la séquence d'ADN (séquence promotrice, séquences « enhancer » et/ou « silencer ») ou au niveau de facteurs de régulation internes (machinerie de transcription) et externes (récepteurs hormonaux). La transcription s'achève lorsque la Pol II atteint un site de terminaison sur l'ADN.

L'ARNpm va d'abord subir des modifications post transcriptionnelles. La terminaison 5' de l'ARNpm va être coiffée d'une 7-méthylguanosine liée par une liaison 5'-5' triphosphate. Cette coiffe stabilise l'ARNm et lui confère une aptitude traductionnelle plus importante. Une séquence poly(A) est ensuite ajoutée à la terminaison 3' par l'intervention d'un complexe multi-protéique. Cette terminaison poly(A) interviendrait dans le transport de l'ARNm du noyau vers le cytoplasme et dans la stabilité de cet ARNm grâce à son association

avec une protéine de liaison (« PolyA Binding Protein », PABP). L'ARNpm va ensuite subir un épissage de ses exons au cours duquel les séquences introniques seront éliminées par clivage de l'intron dans sa partie 5' (GUXXX). Cette partie coupée se termine par une guanosine qui va se lier, par une liaison 2',5'-phosphodiester, à un résidu adénosine spécifique situé 20 à 50 nucléotides en amont du site 3' d'épissage de l'intron, formant ainsi une structure en lasso. Une coupure de l'intron en 3' se produira pour permettre la liaison (épissage) des deux exons situés de part et d'autre de l'intron éliminé. L'excision des introns fait intervenir des complexes spliceosomes composés de particules ribonucleoprotéiques nucléaires (« small nuclear ribonucleoprotein », snRNP).

Chez la vache et la chèvre, les gènes spécifiant les caséines sont tous situés dans une région de 250 kb du chromosome 6 (cf. Figure I.16).

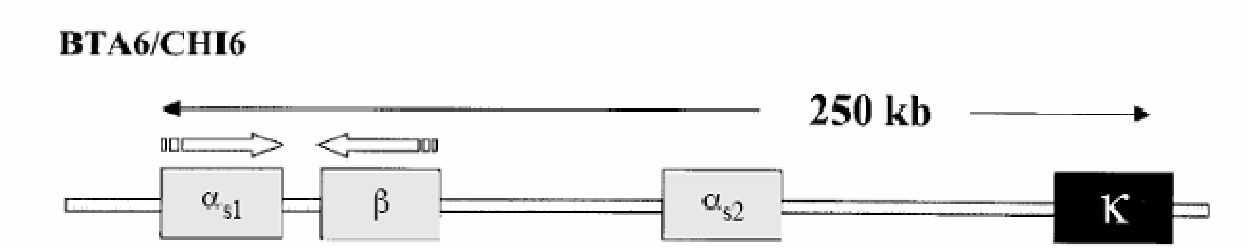


Figure I-16 : Organisation génomique du locus des caséines bovines (Martin *et al.*, 2002).
Les gènes spécifiant les caséines sensibles au calcium sont en gris, le gène spécifiant la caséine κ est en noir.

La Figure I.17 présente la structure des gènes des protéines majeures du lait, s'agissant de l'espèce bovine. Avec ses 19 exons, le gène de la caséine α_{s1} est celui qui en contient le plus. Le gène de la caséine κ comporte 5 exons. Ceux de l'α-lactalbumine et de la β-lactoglobuline contiennent respectivement 4 et 7 exons.

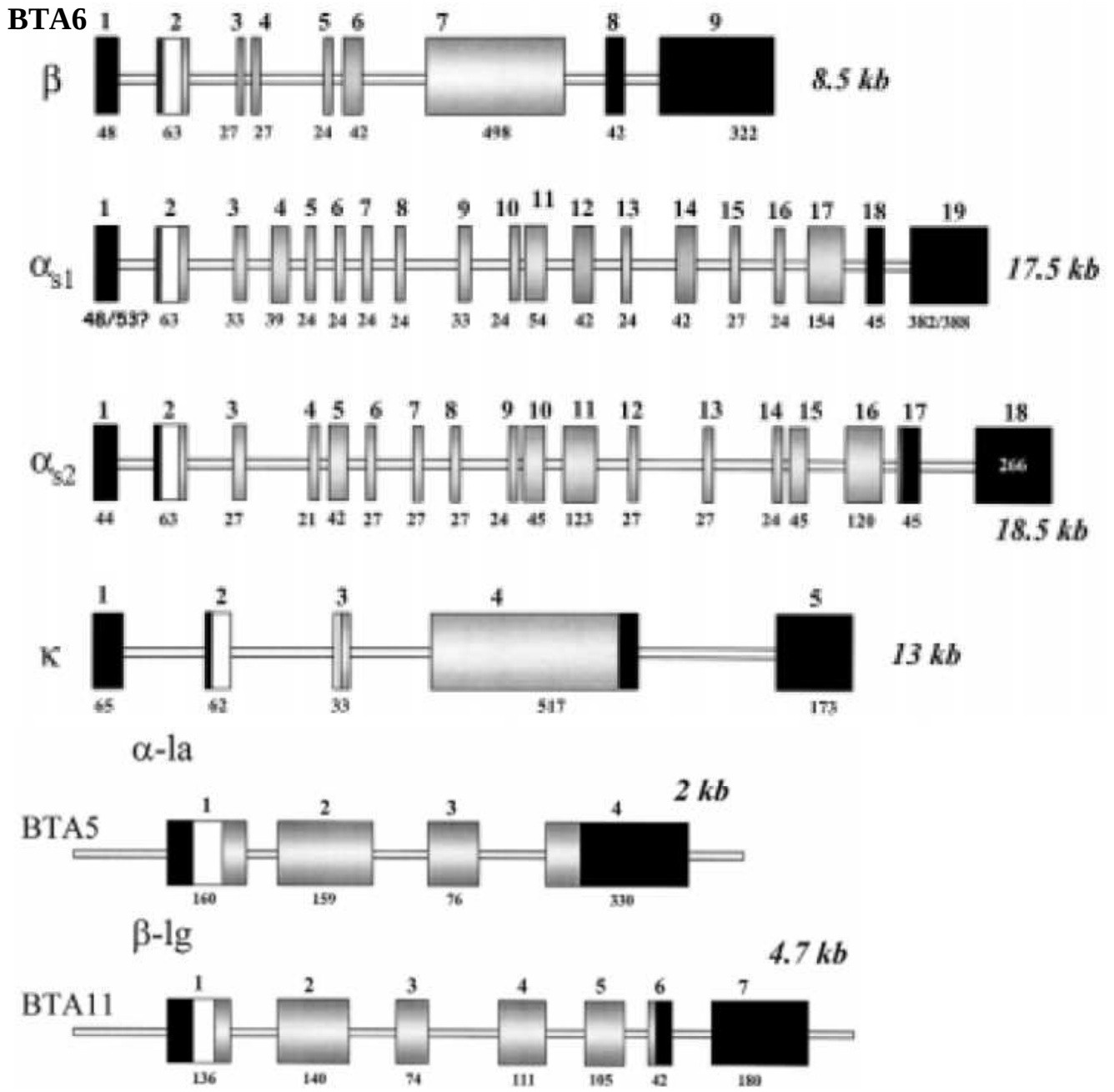


Figure I-17 : Organisation structurale des unités de transcription des caséines, de l'α-lactalbumine et de la β-lactoglobuline bovine (Martin *et al.*, 2002)

Les barres horizontales représentent les introns. Les exons sont représentés par de larges boites noires pour les régions 5' et 3' non traduites, blanches pour la partie de l'exon spécifiant le peptide signal et grises pour les exons spécifiant la protéine mature. La taille des exons est donnée en paires de bases en dessous de chaque boite, le numéro de l'exon est indiqué au dessus. BTA : chromosome bovin, β : caséine β, α_{s1} : caséine α_{s1}, α_{s2} : caséine α_{s2}, κ : caséine κ, α-la : α-lactalbumine, β-lg : β-lactoglobuline.

Une fois les introns excisés, l'ARNm est exporté du noyau vers le cytoplasme à travers des pores situés dans la membrane nucléaire, en réalité le mécanisme est un peu plus complexe avec une maturation des pré-messagers et une traduction concomitantes. Le métabolisme des ARNm est régulé pour partie par des protéines riches en Sérine et Arginine,

les SR protéines (Long & Caceres, 2009). Cet ARNm a une structure contenant une coiffe dans la terminaison 5', suivi d'une séquence non traduite, puis d'un codon AUG correspondant au codon d'initiation de la traduction. Viennent ensuite la séquence traduite qui se termine par un codon STOP (UAG, UAA ou UGA), une séquence 3' UTR (UnTranslated Region) et la queue poly(A) (cf. Figure I.18).

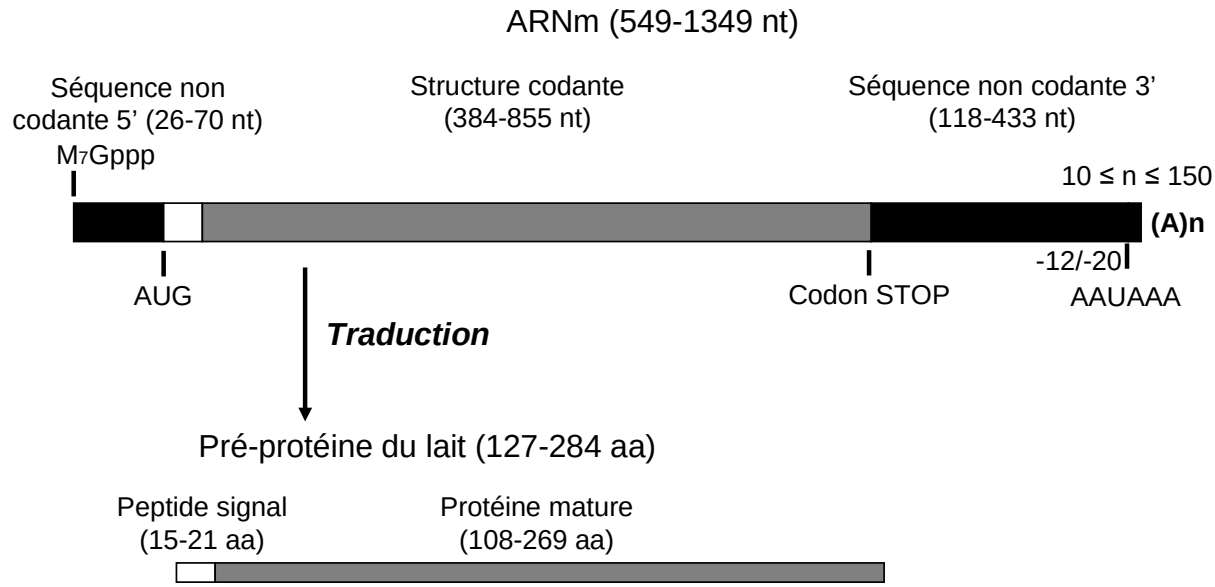


Figure I-18 : Représentation schématique d'un ARNm du tissu mammaire codant les protéines majeures du lait : les 4 caséines, la β-lactoglobuline et l'α-lactalbumine (Mepham *et al.*, 1992).
Les nombres indiquent les tailles extrêmes des ARNm, des protéines et des régions constitutives de 8 espèces étudiées. Les exons sont représentés par de larges boites noires pour les régions 5' et 3' non transcrites, blanches pour la partie du transcrit spécifiant le peptide signal et grises pour les séquences exoniques spécifiant la protéine mature. nt = nucléotide, aa = acide aminé.

➢ De l'ARNm à la protéine mature

L'ARNm est ensuite traduit dans le cytosol, au niveau des ribosomes, en une séquence d'AA par suivi du code génétique. La régulation de cette étape dépend essentiellement de l'accessibilité de la coiffe en 5' et du nombre de structures secondaires dans la région 5' non transcrite et dans les régions entourant le codon d'initiation de la traduction, AUG.

La traduction de l'ARNm en protéine est divisée en trois étapes : l'initiation, l'élongation et la terminaison. Ces trois étapes nécessitent la présence de facteurs protéiques, d'ARN de transfert (ARNt), de ribosomes et d'énergie fournie par la guanosine triphosphate (GTP). Pour plus de détails on pourra se reporter à l'excellente revue de Suzuki, 1995.

Le réticulum endoplasmique (RE) est le « sas de sortie » pour les protéines qui vont être sécrétées telles que les protéines du lait. Une fois dans le RE, les protéines passent transversalement dans la cis, médian puis trans citerne de l'appareil de Golgi avant d'atteindre le réseau trans-Golgien (trans-Golgi network, TGN) à partir duquel elles sont réparties dans des vésicules de sécrétion. L'entrée dans le réticulum se fait grâce à un peptide signal situé en début de la chaîne polypeptidique. Ce peptide signal est reconnu par une particule de reconnaissance du signal (ou « Signal Recognition Particle », SRP) qui est un polypeptide d'un complexe ribonucléoprotéique. La SRP interagit avec un récepteur hétérodimérique de la membrane du RE qui permet la liaison du ribosome avec des ribophorines. La SRP libère la séquence signal qui va interagir avec une protéine N-glycosylée de la membrane du RE appelée récepteur de la séquence signal (« Signal-Sequence Receptor », SSR). Ce SSR est un composant possible pour le transfert dans la lumière du RE de la chaîne polypeptidique croissante (Mepham *et al.*, 1992; Vilotte *et al.*, 2003).

Avant d'être sécrétée, la chaîne polypeptidique va subir des modifications co- ou post-traductionnelles. Une fois passé dans la lumière du RE, le peptide signal est clivé (Mepham *et al.*, 1982). Puis il y a formation de ponts disulfures et repliements de la protéine, orchestrés par des protéines chaperonnes (PDI, BiP/GRP78, Erp77, calnexine et calréticuline). Dans ce compartiment cellulaire, l'α-lactalbumine est également N-glycosylée (Vilotte *et al.*, 2003). Les protéines sont exportées du RE vers l'appareil de Golgi via un système vésiculaire de transport. Dans ce compartiment la caséine κ est O-glycosylée sur des résidus Thréonyle (Thr) et les caséines sont phosphorylées. La phosphorylation des caséines se fait par la reconnaissance enzymatique d'un triplet d'acides aminés –Ser/Thr-X-B- avec X correspondant à n'importe quel AA et B un AA acide (Glu, Asp ou Ser-P) (Mepham *et al.*, 1992). Cette étape est cruciale pour la liaison du calcium aux caséines et pour la formation des micelles de caséines (Vilotte *et al.*, 2003). Le site de formation des micelles n'est pas encore clair, l'agrégation des caséines débuterait dans l'appareil de Golgi et se poursuivrait dans les vésicules de sécrétion. Le nombre de micelles par vésicule de sécrétion varie de 1 à plus de 10, le nombre augmente de façon concomitante avec la quantité de lactose contenue dans la vésicule (Vilotte *et al.*, 2003).

Les protéines du lait sont sécrétées dans la lumière de l'acinus par exocytose des vésicules de sécrétion, consistant à une fusion membranaire et un déversement du contenu des vésicules dans la lumière alvéolaire.

I.3.2.4. Régulation de la synthèse et de la sécrétion des protéines du lait

Il existe plusieurs niveaux de régulation de la synthèse des protéines, la transcription, la stabilité des ARNm, la traduction de ces ARNm en protéines et la sécrétion. Ces mécanismes s'appliquent aussi aux protéines du lait.

La régulation de la transcription est sous contrôle d'un complexe multi-hormonal et de facteurs de croissance locaux. Les hormones lactogènes telles que l'insuline, la prolactine et les glucocorticoïdes activent la transcription de la plupart des gènes des protéines du lait, alors que les autres hormones telles que la progestérone inhibe cette activation. La transcription est également régulée par les composants de la membrane basale. Ainsi, la laminine induirait l'expression de l'α-lactalbumine et de la caséines α_{s1} (Vilotte *et al.*, 2003). Aux *loci* des gènes spécifiant les lactoprotéines il existe des motifs de séquences nucléotidiques très conservées appelés éléments de régulation *cis*. Ces séquences peuvent être reconnus par des facteurs de transcription tels que OCT-1, NF-1, C/EBP, STAT5, GR, Ets-1 et YY1 (Vilotte *et al.*, 2003).

La traduction serait contrôlée par l'insuline et la prolactine qui interagiraient en synergie pour augmenter le niveau de protéines synthétisées (Vilotte *et al.*, 2003).

La sécrétion des protéines serait régulée par l'ocytocine et la prolactine. L'ocytocine augmenterait le transport entre le RE et l'appareil de Golgi des protéines nouvellement synthétisées et favoriserait la circulation de vésicules de sécrétion vers la membrane apicale. La prolactine agirait sur les dernières étapes du processus d'exocytose (Ollivier-Bousquet, 1993).

I.3.3. Les lipides du lait

I.3.3.1. Composition et structure des lipides du lait

Les lipides du lait sont principalement constitués de glycérides (~99% des lipides totaux) qui se répartissent en triglycérides (pour 98%), diglycérides (0,3%) et monoglycérides (moins de 0,1%) Le reste des lipides du lait se compose d'acides gras libres (0,02%), de phospholipides (0,8%), de cérébrosides (0,1%), de stérols (0,35%) et de traces de caroténoïdes (cf. Tableau I.4).

Tableau I-4 : Composition de la matière grasse laitière et distribution des différentes classes de lipides[1]

Classes de lipides	Teneur dans le lait (%)	Distribution (%) dans		
		cœur du globule gras	membrane du globule gras	lactosérum
Glycérides	*98,6*			
Triglycérides	98,3	~100	+	+
Diglycérides	0,3	(90)	(10)	?
Monoglycérides	0,03	+	+	+
Acides gras libres	*0,02*	*75*	+	*25*
Phospholipides	*0,8*	-	*65*	*35*
Lécithine	0,26			
Céphaline	0,28			
Phosphatidylinositol	0,04			
Plasmalogènes	0,02			
Sphingomyéline	0,16			
Cérébrosides	*0,1*	*?*	*70*	*30*
Sterols	*0,35*	*75*	*10*	*15*
Cholestérol	0,3			
Cholestérol esters	0,05			
Caroténoïdes	*0,002*	*(95)*	*(5)*	+

[1]Mulder & Walstra, 1974

La matière grasse du lait est sécrétée sous forme de gouttelettes lipidiques, appelées globules gras, entourées par une membrane complexe, la membrane du globule gras, qui préserve leur intégrité et les rend compatibles avec leur environnement aqueux (Keenan & Dylewski, 1995; Mulder & Walstra, 1974). Les triglycérides sont majoritairement positionnés au cœur lipidique des globules gras. Ils se répartissent de la manière suivante : les triglycérides contenant des acides gras à bas point de fusion (liquide à température ambiante, acides gras insaturés) sont situés au centre de la goutte et les triglycérides contenant des acides gras à haut point de fusion (acides gras long et saturés) sont localisés sur la couche externe de la goutte (Pointurier & Adda, 1969). La membrane du globule gras est formée de trois couches de phospholipides : une monocouche interne provenant de la membrane du réticulum endoplasmique et une bicouche externe dérivant de la membrane plasmique de la CEM. Des protéines, des lipides et des polysaccharides sont ancrés dans cette membrane (Mulder & Walstra, 1974). La taille du globule gras du lait de vache oscille entre 0,2 et 20 µm avec une taille moyenne de 4 µm (Briard *et al.*, 2003; Michalski *et al.*, 2001; Pougheon & Goursaud, 2001).

I.3.3.2. Synthèse des lipides du lait

➢ Synthèse des triglycérides

Les lipides du lait étant essentiellement constitués de triglycérides, la voie de synthèse des triglycérides (TG) représente la voie de synthèse majeure des lipides du lait. Les TG sont

des molécules insolubles dans l'eau, formées d'une molécule de glycérol, une courte chaîne carbonée, dont les trois fonctions alcools sont estérifiés par trois acides gras (AG) (Chrétien, 1995; Mann & Skeaff, 2002). Les acides gras présents sur la molécule de glycérol proviennent de plusieurs sources : ils sont soit synthétisés *de novo* dans la CEM (chaînes de 4 à 16 atomes de carbones), soit captés dans la circulation sanguine par la CEM pour les chaînes de plus de 16 atomes de carbone (Gerson *et al.*, 1967; Hawke & Taylor, 1995).

- **Synthèse des acides gras dans la cellule épithéliale mammaire**

Chez les ruminants, la synthèse des acides gras s'effectue à partir d'acétate ou dans une moindre mesure de β-hydroxybutyrate sanguin, issus de la fermentation des glucides dans le rumen (cf. Figure I.19). Elle nécessite le NADPH comme substrat réducteur. Chez les monogastriques, la synthèse *de novo* des acides gras s'effectue à partir de glucose. Le NADPH provient du métabolisme de l'acétate (dans la mitochondrie) ou du glucose, via la voie hexose monophosphate dans le cytosol (voie 1C sur la Figure I.20).

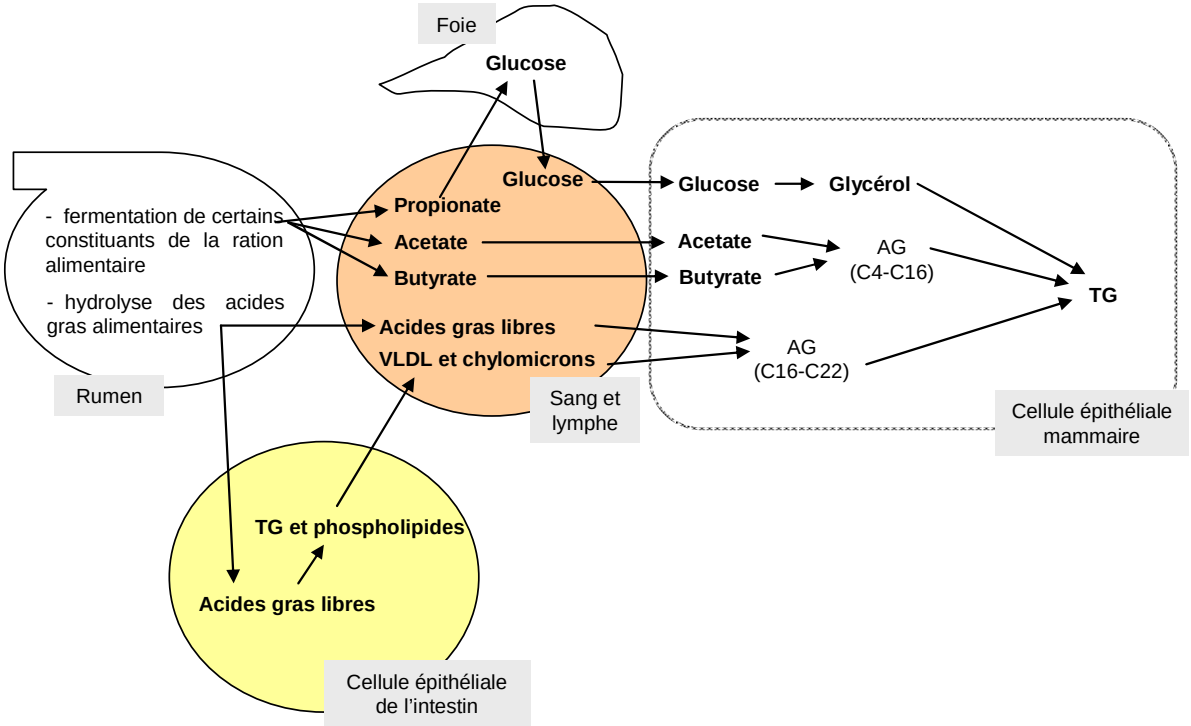


Figure I-19 : Origine des substrats utilisés par la cellule épithéliale mammaire (CEM) pour la synthèse des lipides du lait (d'après Doreau & Ferlay, 1994).

Le lieu de synthèse des acides gras est le cytosol de la CEM. Au cours de la première étape (1A de la Figure I.20), l'acétate est transformé en acétyl-CoA grâce à une acétyl CoA

synthétase (ACS). L'acétyl CoA est ensuite converti en malonyl-CoA, cette réaction est catalysée par l'acétyl CoA carboxylase (ACC, 1B de la Figure I.20).

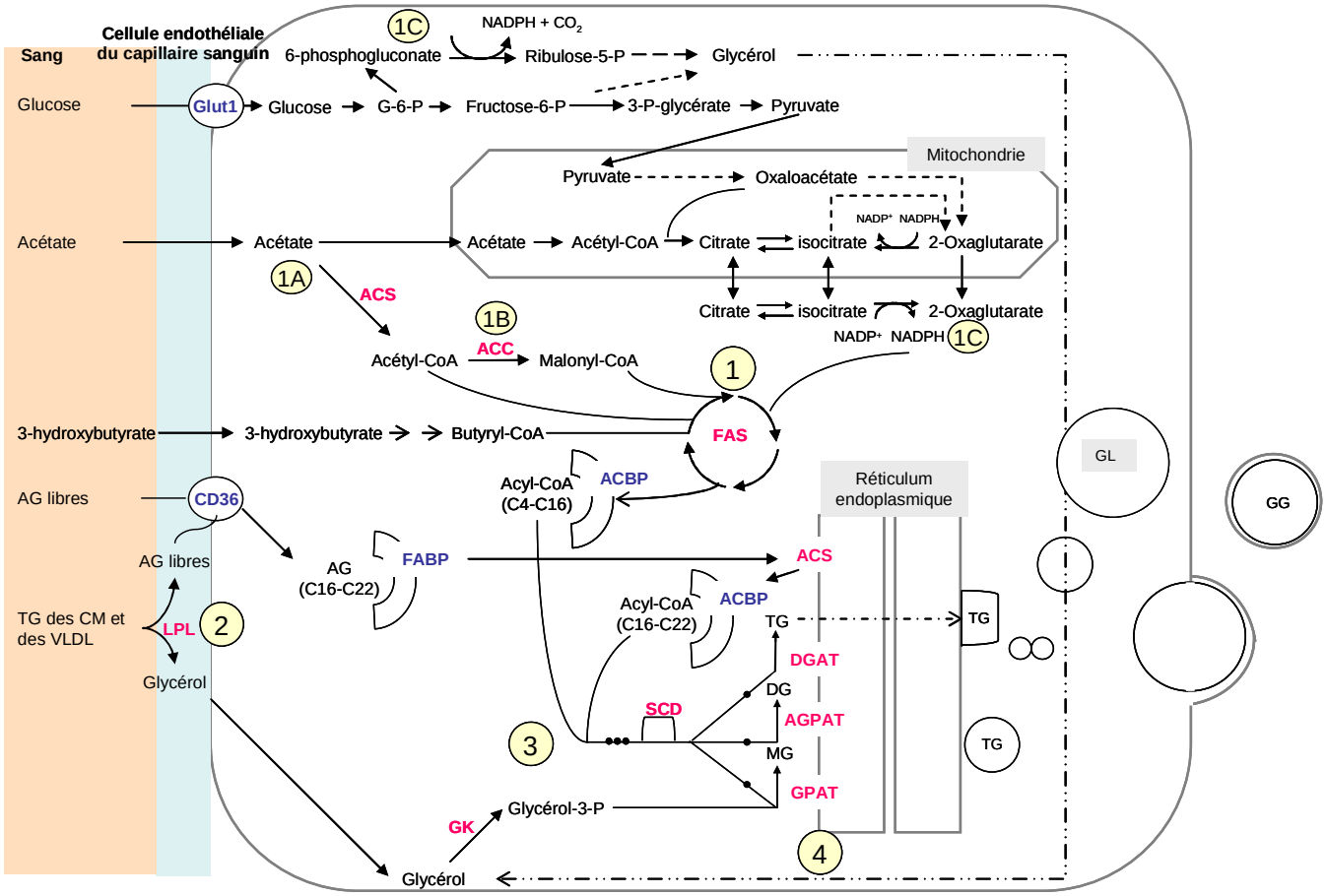


Figure I-20 : Synthèse des acides gras et des triglycérides dans la cellule épithéliale mammaire des ruminants (d'après Bernard *et al.*, 2008 ; Hawke & Taylor, 1995).
1 = voie de synthèse des acides gras ; 2 = captation des acides gras d'origine sanguine ; 3 = désaturation des acides gras ; 4 = synthèse des triglycérides. ACBP : « acyl-CoA binding protein » ; ACC : acétyl-CoA carboxylase ; ACS : acyl-CoA synthétase ; AG : acides gras ; AGPAT : acylglycérol-3-phosphate acyltransférase ; CD36 : « cluster of differentiation 36 » ; CM : chylomicron ; CO2 : dioxyde de carbone ; DG : diacylglycérol ; DGAT : diacylglycérol acyltransférase ; FABP : « fatty acid binding protein » ; FAS : « fatty acid synthase » ; Fructose-6-P : Fructose-6-phosphate ; GG : globule gras ; GK : glycérol kinase ; GL : gouttelette lipidique ; GPAT : glycérol-3-phosphate acyltransférase ; G-6-P : glucose-6-phosphate ; LPL : LipoProtéine Lipase ; MG : monoacylglycérol ; NADP+/NADPH : nicotinamide adénine dinucléotide phosphate ; Ribulose-5-P : Ribulose-5-phosphate ; SCD : « Stearoyl-CoA Desaturase » ; TG : triglycérides ; VLDL : « very low density lipoprotein » ; 3-P-glycérate : 3-phospho-glycérate.

L'initiation de la synthèse des acides gras se fait par la condensation d'un acétyl-CoA avec un malonyl-CoA, cette réaction est catalysée par une synthétase d'acide gras (« fatty acid synthase » ou FAS). L'élongation de la chaîne carbonée se fait ensuite de façon cyclique par ajout de deux carbones venant d'une molécule de malonyl-CoA. La différence de tailles des chaînes carbonées des AG du lait vient de l'interruption du processus d'élongation. Il semble que ce processus d'interruption de l'élongation soit spécifique aux cellules des tissus

mammaires (Hawke & Taylor, 1995). Dans les autres tissus, la réaction d'élongation ne s'achève que lorsque l'acide gras à 16 atomes de carbone (acide palmitique) a été produit. Ces réactions sont catalysées par un complexe enzymatique appelé FAS qui contient sept sites catalytiques (responsable de l'élongation) et un site protéique permettant de lier l'acide gras en cours de synthèse. Ce site protéique est clivé en fin de réaction par une thioesterase afin de libérer l'acide gras. L'acide gras est souvent libéré dans le cytosol sous la forme d'acyl-CoA. Il circule probablement lié à une protéine de liaison (Acyl CoA Binding Protein, ACBP). Le complexe ACBP-Acyl-CoA est un substrat pour les enzymes lipogéniques telles que les synthétases de TG (Mikkelsen & Knudsen, 1987).

La régulation de la synthèse des acides gras dépendrait de la régulation de certaines enzymes limitantes telles que la FAS, l'isocitrate-NADP déshydrogénase et l'ACC. La synthèse des acides gras du lait n'intervient qu'au déclenchement de la sécrétion lactée, lors de la parturition (Mellenberger & Bauman, 1974; Mellenberger *et al.*, 1973). La voie de synthèse des acides gras du lait est donc régulée par les hormones lactogènes à savoir la prolactine, l'insuline et la cortisone (Hawke & Taylor, 1995).

- **Captation des acides gras à longues chaînes par la cellule épithéliale mammaire**

Les acides gras du lait à chaîne longue (de 16 à 22 atomes de carbone) ne sont pas néosynthétisés dans la glande mammaire et ne proviennent pas de l'élongation d'acides gras plus courts mais sont captés à partir du sang par la CEM (voie 2 de la Figure I.20).

Chez les ruminants, ces AG sont présentés à la CEM soit sous forme de TG intégrés dans des lipoprotéines (« Very Low Density Lipoprotein », VLDL ou « Low Density Lipoprotein », LDL) ou des chylomicrons (CM) soit sous forme d'acides gras non estérifiés (AG libres) liés à la sérum albumine. Ces acides gras proviennent de l'alimentation de l'animal (cf. Figure I.19) ou de la mobilisation à partir des dépôts adipeux (Hawke & Taylor, 1995). La captation des acides gras contenus dans les TG circulants se fait après hydrolyse des TG par la lipoprotéine lipase (LPL). Cette lipase est liée à la membrane plasmique luminale des cellules endothéliales des capillaires sanguins. Les acides gras libérés par la lipoprotéine lipase seraient transportés par des mouvements latéraux à travers la membrane luminale de la cellule endothéliale et de la membrane basale de la CEM (Scow *et al.*, 1979). Une autre hypothèse suppose que les acides gras libres soient transportés à travers la membrane plasmique de la CEM grâce au cluster de différenciation 36 (« cluster of

differentiation », CD36), avant d'être liés à une protéine de liaison aux acides gras (« Fatty Acid Binding Protein », FABP), localisée dans le cytoplasme (Rasmussen *et al.*, 1998).

Les acides gras captés du sang sont ensuite modifiés en acyl-CoA par une acyl-CoA synthétase (ACS) longue ou moyenne chaîne située dans la membrane du réticulum endoplasmique (Hawke & Taylor, 1995). Ils sont ensuite relargués dans le cytosol où ils sont liés à ACBP ou intégrés directement aux réactions de synthèse des TG.

Pendant la lactation, la prolactine entraînerait une suppression de l'activité de la LPL du tissu adipeux et une augmentation de la quantité de cette lipase dans le tissu mammaire (pour revue, Hawke & Taylor, 1995).

- **Modification des acides gras pour la synthèse des triglycérides du lait**

Dans la CEM totalement différenciée, s'exerce une forte activité de désaturation liée à l'activité de la delta-9-désaturase ou stéaroyl-CoA désaturase (SCD, voie 3 de la Figure I.20). La SCD désature les AG à chaîne carbonée longue et moyenne (supérieure à 10 atomes de carbone) avec un effet plus marqué pour les acides gras de 10 à 16 atomes de carbone (Schennink *et al.*, 2008). Dans la CEM, la SCD introduit une double liaison entre les carbones 9 et 10 (cf. Figure I.21).

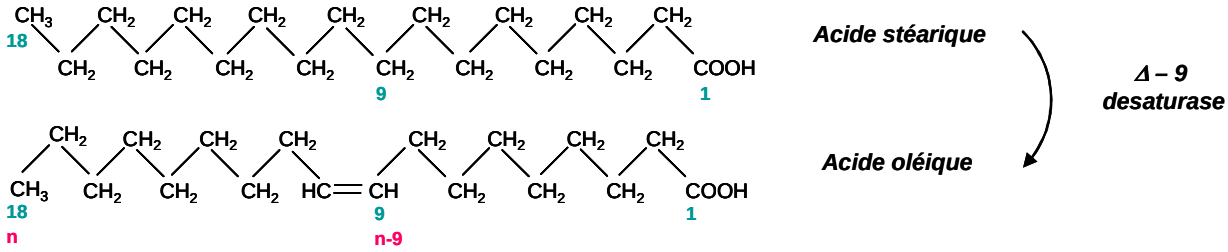


Figure I-21 : Désaturation de l'acide stéarique (C18:0) en acide oléique (C18:1) par la delta-9 désaturase

Les autres désaturations observées sur les acides gras du lait, telles que la désaturation au niveau du 11ème carbone sur l'acide trans-vaccénique, sont des désaturations issues des AG alimentaires polyinsaturés qui ont été partiellement hydrogénés par les microorganismes du rumen. Les substrats préférentiels de la SCD sont les acyl-CoA apportés par la FABP3. Les acyl-CoA désaturés sont ensuite transportés par la FABP4 (Bionaz & Loor, 2008) puis incorporés dans la matière grasse laitière.

La désaturation des acides gras longs de plus de 20 atomes de carbone est possible grâce à la présence de désaturases d'acides gras (« fatty acid desaturase », FADS) qui ajoutent des doubles liaisons au niveau des positions delta-5 et delta-6 des acides gras polyinsaturés. C'est

cette enzyme qui catalyse la synthèse des acides arachidoniques (C20:4, *n-6*), eicosapentaénoiques (C20:5, *n-3*) et docosahexaénoiques (C22:6, *n-3*).

- **Biosynthèse des triglycérides dans la cellule épithéliale mammaire**

La biosynthèse des TG dans la cellule épithéliale mammaire (voie 4 de la Figure I.20) utilise préférentiellement le glycérol-3-phosphate (G-3-P, cf. Figure 1.22), qui provient soit du glycérol soit du dihydroxyactéone phosphate. Le G-3-P est synthétisé dans la CEM à partir du glycérol via l'action d'une glycérol kinase ou à partir du glucose via la glycolyse ou la voie hexose monophosphate. La proportion de glycérol dérivée du glucose dans la CEM des ruminants est de 50-60% contre 80% chez la lapine, le reste du glycérol vient de l'hydrolyse des TG du sang par la lipoprotéine lipase (Hawke & Taylor, 1995). L'autre source de G-3-P pour la synthèse des TG est l'acylation du dihydroxyacétone phosphate (DHAP) dérivé du métabolisme glucidique. Cette source est 15 fois moins utilisée dans la CEM que le G-3-P (Hawke & Taylor, 1995).

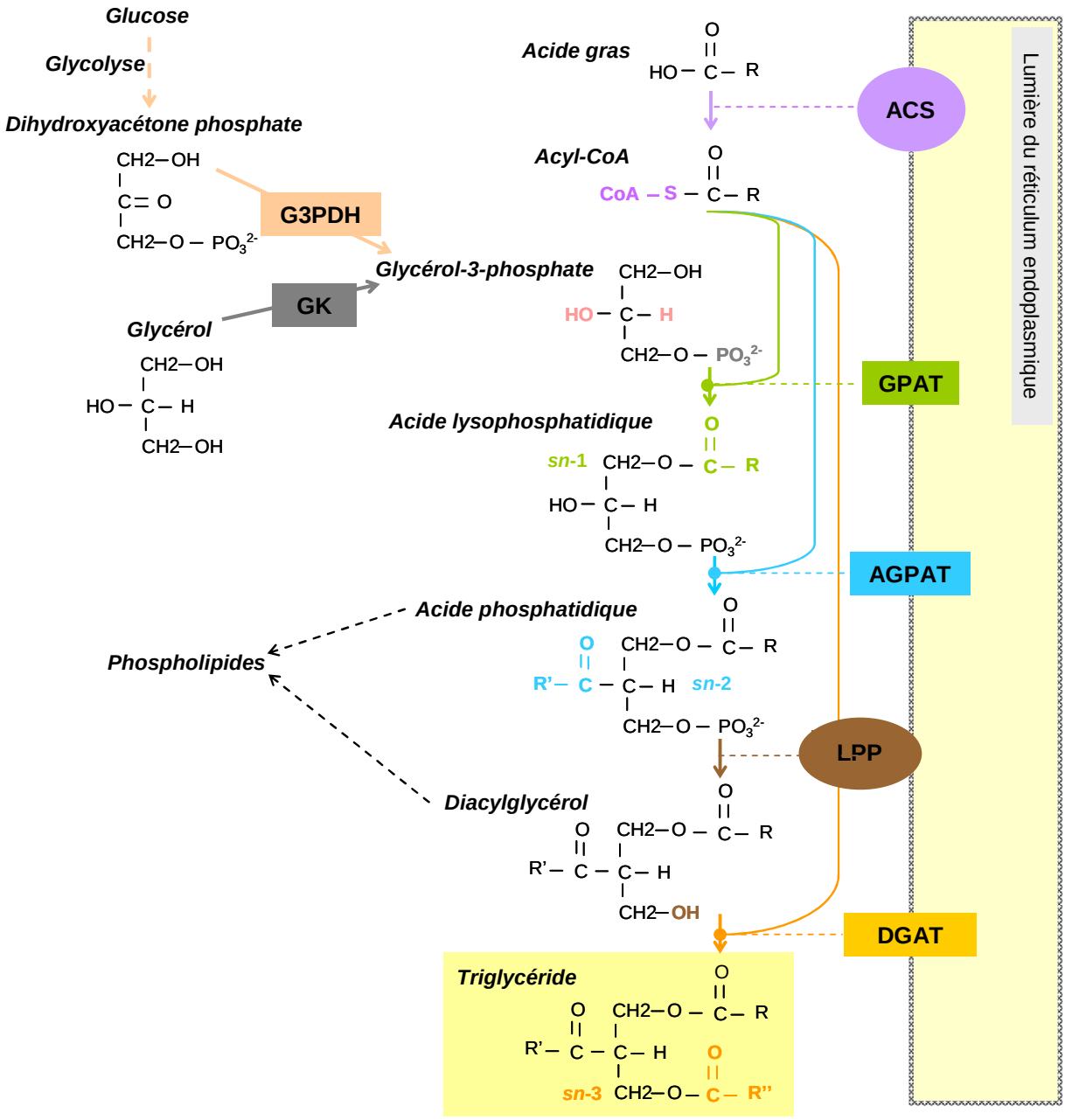


Figure I-22 : Synthèse des triglycérides dans la cellule épithéliale mammaire (adapté de Bell *et al.*, 1981; Ollier, 2007).
ACS : Acyl-CoA synthétase ; AGPAT : acylglycérol-3-phosphate acyltransférase ; DGAT : diacylglycérol acyltransférase ; G3PDH : glycérol-3-phosphate déshydrogénase ; GK : glycérol kinase ; GPAT : glycérol-3-phosphate acyltransférase ; LPP : acide phosphatidique phosphatase.

La voie de synthèse des triglycérides (cf. Figure 1.22) utilise les substrats G-3-P et les acyl-CoA et les enzymes glycérol-3-phosphate acyltransférase (GPAT), acylglycérol-3-phosphate acyltransférase (AGPAT ou acide lysophosphatidique acyltransférase, LPAAT) et la diacylglycérol acyltransférase (DGAT). Ces enzymes apparaissent étroitement liées à la membrane du réticulum endoplasmique (Coleman *et al.*, 2000) et sont appelées plus communément synthétases des TG. Le glycérol contient trois fonctions alcool pouvant

accueillir chacune un acyl-CoA. Ces fonctions alcool portent, dans l'ordre d'estérification par les synthétases des TG, les positions *sn-1*, *sn-2* et *sn-3*, les positions *sn-1* et *sn-3* étant les plus extérieures. La GPAT catalyse l'ajout du premier acide gras en position *sn-1* d'une molécule de glycérol. L'AGPAT participe à l'estérification du deuxième acide gras en position 2 du monoglycérol. DGAT catalyse l'addition du dernier acide gras sur une molécule de diacylglycérol. Une autre enzyme nécessaire à la synthèse des TG est l'acide phosphatidique phosphatase (LPP), c'est une protéine cytosolique qui hydrolyse l'acide phosphatidique pour former le diacylglycérol. Elle est active dans cette réaction lorsqu'elle est associée à la membrane du réticulum endoplasmique (Coleman *et al.*, 2000) et est indispensable pour la synthèse des TG puisque le substrat de DGAT est le diacylglycérol et non l'acide phosphatidique. Il est aussi suggéré dans la littérature que la transformation de l'acide phosphatidique en diacylglycérol soit catalysée par la lipine 1 (Bionaz & Loor, 2008). Plusieurs hypothèses sont possibles et non exclusives pour l'utilisation des acides gras à chaînes carbonées courtes et moyennes au niveau des synthétases de TG. Soit ces acides gras sont directement transmis par la FAS aux synthétases pour la synthèse des glycérolipides (Hawke & Taylor, 1995), soit les acides gras sont présentés aux synthétases de TG sous la forme de substrat complexe ACBP-Acyl-CoA (Mikkelsen & Knudsen, 1987).

La spécificité de substrat des enzymes intervenant dans la synthèse des TG explique pour partie la distribution des acides gras sur les trois positions du glycérol. Les grandes proportions (56-62%) des acides gras estérifiés aux positions *sn-1* et *sn-2* sur le glycérol sont des acides gras de chaîne moyenne, longue et saturée (de C10:0 à C18:0). Le C16:0 est également distribué sur les deux positions (Bernard *et al.*, 2008), les C8:0 à C14:0 sont préférentiellement en position *sn-2* et le C18:0 est préférentiellement positionné en *sn-1*. Les acides gras estérifiés en position *sn-3* sont des acides gras à chaîne courte (C4:0, C6:0, C8:0) et l'acide oléique (C18:1, *cis9*).

Les enzymes de la synthèse des lipides du lait voient leur activité accrue par la présence d'insuline (Coleman *et al.*, 2000). Le glucagon diminue le nombre de transcrits de la GPAT (Coleman *et al.*, 2000).

➢ La biosynthèse des autres lipides du lait

A part les acides gras non libres, les autres lipides du lait sont essentiellement des constituants de la membrane du globule gras.

♦ **Synthèse des glycéroplipides.**

La plupart des phospholipides présents dans la membrane des globules gras sont synthétisés *de novo* dans la cellule épithéliale mammaire à partir des produits de l'hydrolyse pendant leur absorption des phospholipides présents dans le sérum. La phosphatidylcholine, le phospholipide le plus représenté dans le lait, et la phosphatidyléthanolamine sont synthétisés dans la glande mammaire à partir du diacylglycérol (cf. Figure I.23).

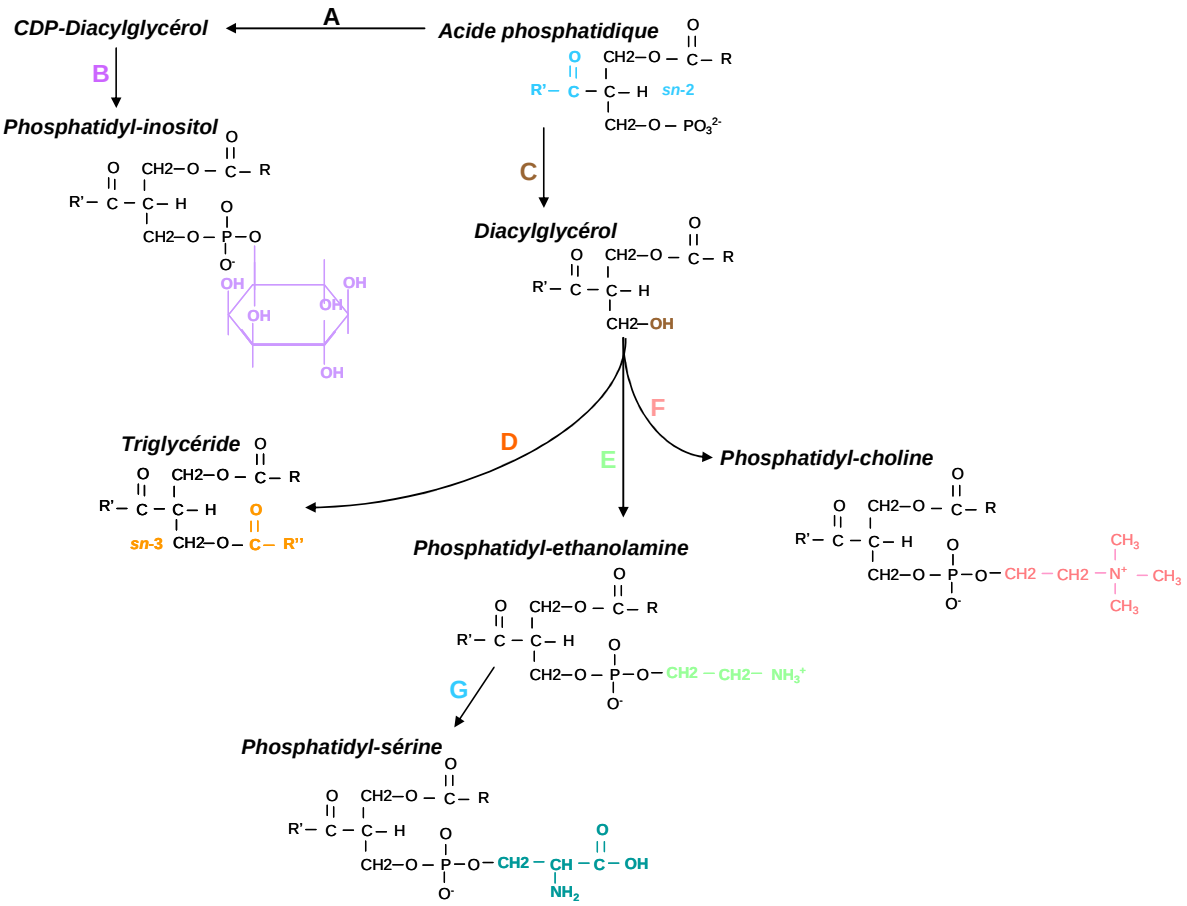


Figure I-23 : Enzymes intervenant dans la synthèse des phospholipides du lait (adapté de Bell *et al.*, 1981). A : acide phosphatidique cytidyltransférase (CDP-diacylglycérol synthétase), B : phosphatidylinositol synthétase, C : acide phosphatidique phosphatase, D : diacylglycérol acyltransférase, E : diacylglycérol éthanolamine phosphotransférase, F : diacylglycérol cholinephosphotransférase, G : phosphatidyléthanolamine serinetransférase.

Ces synthèses font intervenir la diacylglycérol choline phosphotransferase et la diacylglycérol éthanolamine phosphotransferase qui utilisent respectivement la CDP-choline et la CDP-éthanolamine comme substrat (Bell *et al.*, 1981). Ces enzymes sont probablement ancrées dans la membrane du réticulum endoplasmique. La phosphatidyl sérine est issue de la carboxylation de la phosphatidyl éthanolamine par la phosphatidyl éthanolamine sérine

transférase. La phosphatidyl inositol est synthétisée à partir de CDP-diacylglycérol grâce à l'action de la phosphatidyl inositol synthétase. La sphingomyéline est synthétisée à partir de céramide lui même synthétisé *de novo* dans la CEM à partir de sérine et d'acide palmitique (C16:0).

- **Le cholestérol**

Chez les ruminants, la part de cholestérol provenant de l'alimentation est très faible, voire nulle, car les régimes sont pauvres en cholestérol et celui-ci est hydrogéné dans le rumen. Le cholestérol présent dans le lait est synthétisé *de novo* à partir d'acétate soit dans le foie, puis passe dans le sang avant d'être capté par la glande mammaire, soit directement dans la cellule épithéliale mammaire (Hawke & Taylor, 1995).

I.3.3.3. Origine et sécrétion du globule gras

Les précurseurs intracellulaires du globule gras prennent leur source dans le réticulum endoplasmique, qui est le lieu de synthèse des triglycérides. Après leur synthèse, les triglycérides s'accumulent entre les deux couches phospholipidiques de la membrane du réticulum endoplasmique, créant ainsi un renflement (cf. Figure I.24). Après avoir grossi, le renflement se détache du réticulum endoplasmique et est libéré dans le cytosol sous la forme de petites gouttelettes lipidiques entourées d'une monocouche de phospholipides. Le mécanisme associé à la formation de ces gouttelettes, notamment l'accumulation de TG et sa libération, reste encore largement hypothétique pour ne pas dire inconnu. Des protéines telles les périlipines, TIP47 et l'adipophiline interviendraient dans ce mécanisme dont la régulation, si ce processus est régulé reste à explorer (Bauman *et al.*, 2006; Heid & Keenan, 2005).

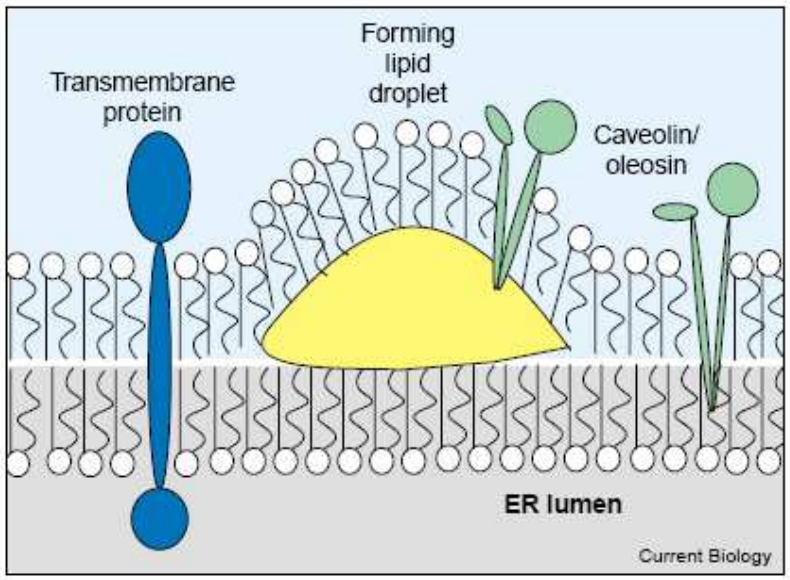


Figure I-24 : Formation de la gouttelette lipidique entre les deux couches phospholipidiques de la membrane du réticulum (Brown, 2001).
En jaune les triglycérides.

Dans le cytosol des cellules épithéliales mammaires, la taille des gouttelettes lipidiques varie de moins de 0,5 µm à plus de 4 µm de diamètre. Les gouttelettes augmenteraient de volume par fusion entre elles, bien que les mécanismes de fusion ne soient pas encore bien élucidés. Les mécanismes de fusion entre gouttelettes existeraient entre les petites gouttelettes. Lorsque la gouttelette aurait atteint un certain volume, la fusion ne serait plus possible. L'augmentation de la taille des gouttelettes lipidiques se ferait également par transfert de lipides du cytosol vers l'intérieur de la gouttelette (Heid & Keenan, 2005). L'augmentation du volume des gouttelettes lipidiques s'intensifie au voisinage de la membrane apicale de la CEM. Les globules gras du lait de vache sont à 80% d'un diamètre inférieur à 1 µm, ce qui laisse supposer qu'une grande majorité des gouttelettes lipidiques ne fusionnent pas avant leur sécrétion.

Les gouttelettes lipidiques migrent de leur lieu de formation, le plus souvent au niveau basal de la cellule, vers la membrane apicale par des mécanismes non entièrement élucidés. Il semble toutefois que le cytosquelette intervienne dans ces mécanismes de migration. Les microtubules, les filaments d'actine et les filaments de prékeratine sont autant de pistes d'étude sur la migration des gouttelettes lipidiques dans la CEM (Heid & Keenan, 2005).
Les gouttelettes lipidiques sont sécrétées au niveau apical par enroulement progressif de la gouttelette par la membrane plasmique jusqu'à son expulsion totale dans la lumière de l'acinus. Un autre mécanisme de sécrétion possible passe par la voie de sécrétion des protéines. Dans ce mécanisme, la gouttelette serait associée à des vésicules de sécrétion et

serait donc expulsée par exocytose dans la lumière de l'acinus (cf. voie C Figure I.2, Mather & Keenan, 1998b). Toutefois, compte tenu de la composition protéique de la membrane du globule gras, qui révèle la présence de protéines telles que la butyrophiline (BTN) et la mucine 1 (MUC1), le premier mécanisme de sécrétion apparaît être dominant. Lors de la sécrétion de la gouttelette lipidique, la xanthine oxydase (XDH) située dans le cytosol se lierait à deux molécules de BTN situées dans la membrane plasmique formant ainsi des dimères de BTN. La XDH et les BTN sont liées par des ponts disulfures. Le complexe BTN-XDH ainsi formé pourrait se lier à l'adipophiline (ou « adipocyte differentiation-related protein », ADRP) située dans la monocouche lipidique entourant la gouttelette lipidique. Ce sont ces interactions entre protéines qui permettraient l'enrobage total de la gouttelette par la membrane plasmique et l'expulsion du globule gras dans le la lumière des acini. Lors de l'expulsion de la gouttelette lipidique, du cytoplasme peut être également enveloppé par la membrane plasmique de la CEM (Heid & Keenan, 2005). Le globule gras mature ensuite dans la citerne de la glande mammaire. Au cours de cette maturation, des échanges se font entre le contenu de la lumière de l'acinus, le cœur et la membrane du globule gras : les éléments liposolubles se déplaceraient essentiellement de la membrane vers le noyau du globule gras, les éléments hydrosolubles se déplaceraient préférentiellement vers le lait, seuls les éléments amphiphiles resteraient ancrés dans la membrane (pour revue, Evers, 2004).

I.3.3.4. Le réseau de gènes impliqués dans la synthèse de la matière grasse laitière

Les travaux présentés dans cette thèse comparent les profils d'expression génique dans le tissu mammaire. Nous avons donc choisi de lister dans cette partie les isoformes des gènes exprimés dans le tissu mammaire des ruminants en lactation, et spécifiant des enzymes impliquées dans la synthèse de la matière grasse laitière. Pour plus de détails et de précisions, le lecteur pourra se référer aux revues de Bernard et al. (2008) et Bionaz et Loor (2008).

Les gènes participant à la captation et au transport des acides gras vers la CEM sont ceux du récepteur des VLDL (*VLDLR*), de la lipoprotéine lipase (*LPL*), des transporteurs d'acide gras (« solute carrier family » 27, *SLC27A6*) et le *CD36*.

Les gènes intervenant dans le transport des acides gras dans la CEM spécifient les protéines de liaison aux acides gras (« Fatty Acid Binding Protein », *FABP3* et *FABP4*) et les protéines de liaison aux acyl-CoA, bien que leur rôle soit moindre (« Acyl-CoA Binding Protein », *ACBP*).

Les gènes associés à l'activation des acides gras pour leur utilisation dans la synthèse des TG spécifient l'acyl-CoA synthétase pour longue chaîne (*ACSL1*) et l'acyl-CoA synthétase pour chaîne courte (*ACSS2* et dans une moindre mesure *ACSS1*).

Les gènes spécifiant des transporteurs d'ATP associés aux membranes sont « ATP-binding cassette, sub-family G (WHITE), member 2 » (*ABCG2*) et l'« ATP-binding cassette, sub-family A, member 1 » (*ABCA1*). La protéine ABCG2 est présente dans la membrane apicale des CEM de souris et joue un rôle actif dans la sécrétion de toxines dans le lait (détoxification du tissu). Elle est aussi présente dans la membrane des globules gras. Son rôle dans les mécanismes de synthèse et de sécrétion du lait est encore mal connu. Elle interviendrait dans la sécrétion de quelques constituants du lait (van Herwaarden *et al.*, 2006). La protéine ABCA1 interviendrait dans l'excrétion du cholestérol par la cellule (Bionaz & Loor, 2008).

Les gènes intervenant dans la synthèse *de novo* des acides gras sont l'acétyl-CoA carboxylase (*ACACA*) et la fatty acid synthase (*FASN*).

La désaturation des acides gras fait intervenir la stéaroyl-CoA désaturase (*SCD1*) pour les acides gras à chaîne carbonée moyenne et longue (entre 10 et 18 atomes de carbone) et les « fatty acid desaturase » (désaturase d'acides gras, *FADS1* et *FADS2*) pour les acides gras à chaîne carbonée très longue ($\geq$ 20 atomes de carbone).

La synthèse des triglycérides est assurée par : la glycérol-3-phosphate acyltransferase, forme mitochondriale (*GPAM*) pour l'estérification du 1er acide gras sur le glycérol, l'1-acylglycerol-3-phosphate-O-acyltransferase 6 (*AGPAT6*) pour l'estérification du 2ème acide gras sur le monoglycérol, l'acide phosphatidique phosphatase (*LPP*) ou la lipine 1 (*LPIN1*) pour l'hydrolyse de l'acide phosphatidique en diacylglycérol, et la diacylglycérol acyltransférase 1 (*DGAT1*) pour l'estérification du 3ème acide gras sur le diacylglycérol.

Les gènes participant à la formation, au transport et la sécrétion des gouttelettes lipidiques spécifient les périlipines (*PLIN*) et l'adipophiline (*ADFP*) qui interviennent dans la formation des gouttelettes au niveau du réticulum endoplasmique et dans leur transport jusqu'à la membrane apicale de la CEM ; la butyrophiline (*BTN1A1*) et la xanthine oxydase (*XDH*) qui interviennent dans les mécanismes de sécrétion de la gouttelette.

La régulation de la synthèse des lipides du lait implique la protéine de liaison à des éléments stérols régulateurs (« stearoyl regulatory element-binding protein 1 », *SREBP1* et dans une moindre mesure *SREBP2*), la protéine de clivage activant la SREBP (« SREBP cleavage activating protein », *SCAP*), le gène activé par l'insuline (« insulin induced gene », *INSIG1* et *INSIG2*), les gènes activés par l'hormone thyroïdienne (« thyroid hormone

responsive », SPOT14, *THRSP*), les récepteurs nucléaires « peroxisome proliferator-activated receptor gamma » (*PPARγ*) et ses co-activateurs (*PPARGC1A*, *PPARGC1B*).
Enfin, les gènes intervenant dans la synthèse des céramides, molécules impliquées dans la signalisation, le cycle cellulaire et la régulation du transfert de protéines du réticulum endoplasmique à l'appareil de Golgi spécifient la « N-acylsphingosine amidohydrolase-like » (*ASAHL*), la « lymphocyte activated gene » (LAG1 homolog, ceramide synthetase 2 ou *LASS2*), le transporteur des céramides entre le RE et le Golgi (« oxysterol binding protein », *OSBPL2* et *OSBPL10*).

Nous avons entrepris d'étudier et de comparer le transcriptome de la glande mammaire dans différents états physiologiques (chapitres II et III). Le transcriptome se définit par tous les ARN présents dans une cellule ou dans un tissu qui vont être traduits en protéines. Il représente un « instantané » de tous les gènes spécifiant des protéines exprimées à un moment donné. Pour étudier le transcriptome, plusieurs techniques existent. Le SAGE (« Serial Analysis of Gene Expression ») qui consiste à séquencer sans *a priori* tous les ARNm extraits d'un tissu ou d'une cellule. Au début des travaux de cette thèse, cette méthode était peu utilisée, notamment en raison des problèmes posés pour l'identification des séquences pour des espèces « orphelines » et du coût des opérations de séquençage. Depuis, avec le développement de nouvelles technologies de séquençage massivement parallèle, on assiste à un regain d'intérêt pour ce type d'approche. La RT-PCR quantitative en temps réel permet de dénombrer les transcrits issus d'un gène donné. Elle est classiquement utilisée pour valider des différentiels d'expression. Avec l'apparition de la technologie TLDA (« TaqMan® Low Density Array ») il est possible d'analyser simultanément 384 gènes de cette façon. Enfin, les puces à ADN permettent la comparaison sur un même support du niveau d'expression de milliers de gènes, voir d'un génome entier. Au début de nos études, cette dernière technique était en pleine émergence, notamment avec le développement de puces à ADN bovine et de modules d'analyse et d'interprétation des données. Il a donc été décidé d'utiliser la technique des puces à ADN « deux couleurs » pour réaliser les travaux. Cette technique sera présentée ci-après.

I.4. Technologie des puces à ADN « deux couleurs » pour l'étude du transcriptome

L'objectif de cette partie est de donner un bref aperçu de la technologie des puces à ADN deux couleurs, utilisée dans le cadre de cette thèse, pour analyser le transcriptome. Pour

plus de détails sur la diversité des microarrays, leurs fabrications et leurs utilisations, le lecteur pourra se référer à l'article de Bogard et al. (2008).

I.4.1. Définition et principe

Une puce à ADN (ou biopuce, *biochips, DNA-microarrays ou microarrays*) est un support (membrane de nylon, lame de verre ou de silicium) sur lequel une collection de fragments d'ADN, de séquence nucléotidique connue, a été fixée de façon ordonnée. Les fragments d'ADN de séquence connue sont appelés des *sondes*. La technique des puces à ADN permet de quantifier le niveau d'expression des gènes, représenté par le nombre de transcrits, dans une cellule ou un tissu donné (foie, intestin, glande mammaire...), à un instant donné du développement ontogénétique (embryon, adulte...) ou dans un état physiopathologique donné (malade/sain, gestation, lactation ...).

Le principe de la puce à ADN repose sur la propriété qu'ont deux molécules d'acide nucléique monocaténaires de s'apparier de façon spécifique, lorsqu'elles présentent un degré de complémentarité suffisant. Concrètement, les ARN totaux sont extraits des cellules et convertis en ADN complémentaires (ADNc) par la technique de transcription inverse. Cette étape de rétro-transcription peut être simultanée ou suivie du marquage par deux fluorochromes tels que les Cyanine 3 (Cy3, verte) et 5 (Cy5, rouge). Les ADNc marqués sont communément appelés *cibles (complexes)*. Une fois marquées ces cibles sont hybridées sur les sondes de la lame de verre. Cette hybridation doit être hautement spécifique entre la cible et la sonde dans les conditions expérimentales utilisées. En système hétérologue (où la sonde et la cible sont de deux espèces différentes), la spécificité peut être légèrement diminuée, l'hybridation peut se faire même s'il existe quelques mismatches.

La technologie des puces à ADN a pour origine une technique de biologie moléculaire appelée Northern Blot, adaptée aux ARN à partir de la technique du Southern Blot développée par Southern en 1975 (revue dans Saluz *et al.*, 2002). Le Northern Blot consiste à séparer dans un gel, par électrophorèse, les ARN extraits d'un tissu dont on veut analyser le profil d'expression génique. Les ARN sont ensuite transférés par adsorption sur une membrane de nylon. Les ARN fixés sur la membrane sont révélés par hybridation d'une sonde nucléotidique spécifique d'un gène donné, marquée radioactivement. Si l'ARN de ce gène est présent sur la membrane, une bande est alors révélée. La technique des puces à ADN procède de façon semblable en multipliant le nombre de sondes géniques révélées en une seule hybridation.

I.4.2. Les différents types de « puces à ADN »

Les différents types de réseaux existant se définissent par la nature du support, le type de sondes déposées, la densité du dépôt et le type de marquage des cibles.

Depuis sa création, la technologie des puces à ADN a connu deux évolutions majeures (Saluz *et al.*, 2002). La première évolution a porté sur la nature du support utilisé. Il y a eu miniaturisation du support et passage à des lames de verre. La seconde évolution a concerné les sondes présentes sur le support. Vers 1994 est apparu, avec les puces Affymétrix, la synthèse *in situ* des séquences oligonucléotidiques, directement sur le support. Ce nouveau mode de fabrication est venu concurrencer le dépôt (spotting) et a permis d'augmenter la densité de gènes présents sur le support (Saluz *et al.*, 2002).

Les sondes spottées sur les lames sont soit des produits de PCR ou des séquences oligonucléotidiques synthétisées chimiquement. Ces sondes sont de séquence connue et peuvent être identifiées comme une séquence exprimée, mais non annotée (« Expressed Sequence Tag », EST) ou correspondre à un gène parfaitement caractérisé.

Les macro-réseaux ou « macroarrays » sont le plus souvent des membranes de nylon sur lesquelles ont été déposées des produits de PCR issus de l'amplification de banques d'ADNc. Les cibles complexes hybridées sur ces réseaux sont marquées radioactivement. Il n'y a qu'un seul échantillon hybridé par membrane, mais une membrane peut être deshybridée et resservir dans plusieurs expériences. La densité des sondes déposées sur la membrane est d'environ 100 sondes/cm².

Les micro-réseaux ou « microarrays » sont essentiellement des lames de verre sur lesquelles ont été déposés des produits de PCR de banque d'ADNc ou des oligonucléotides. Sur ce type de réseaux, les oligonucléotides ont généralement une taille comprise entre 40 et 80 nucléotides. Les cibles complexes hybridées sur ces réseaux sont classiquement marquées par des fluorochromes et une à trois conditions expérimentales (généralement deux) peuvent être hybridées sur une même lame. La densité de dépôt des sondes sur la lame est de 2 000 sondes/cm².

Dans le cas de sondes synthétisés *in situ*, la densité des sondes sur le micro-réseau peut augmenter pour passer de 12 000 sondes/cm² (puces Agilent) à 162 000 sondes/cm² (puces Nimblegen).

Dans ce travail de thèse trois types de puces à oligonucléotides bovins ont été utilisés, suivant leur disponibilité (cf. Tableau I.5).

Tableau I-5 : Caractéristiques des trois puces à ADN utilisées dans ce travail de thèse

	Lame Opéron	**Lame du CRB**	**Lame Agilent**
Taille de oligonucléotides	70mer	70mer	60mer
Méthode de fabrication de la lame	Spotting	Spotting	Synthèse in situ
Lieu de fabrication	Génopôle de Toulouse	CRB-GADIE* Jouy-en-Josas	Agilent USA
Nombre de réseaux par lame	1	1	4
Nombre de spots par réseau	19 200	23 232	44 407
Nombre de set par réseau	2	2	1
Nombre de séquences oligonucléotidiques (sondes) différentes par réseau	8 379	21 586	21 475
Nombre de gènes différents par set ayant une annotation acceptée par IPA	6 656	9 973	13 746
Nom des sets	**Clermont**	**Opéron**	**Illumina**	**Agilent**
Nombre de séquences oligonucléotidiques (sondes) différentes par set	*50*	*8 329*	*13 257*	*21 475*
Nombre de gènes différents par set ayant une annotation acceptée par IPA	*46*	*6 629*	*8 058*	*13 746*
Type de marquage	Direct Pronto (Promega, France)	Indirect (Invitrogen, France) ou Amplification et marquage direct	Amplification et marquage direct (Agilent Technologies)	
Pré-hybridation	Universal microarray hybridation kit (Promega, France)	Solution de BSA	Pas de pré-traitement	
Hybridation	Chambres Corning	Four à hybridation Agilent (Agilent Technologies)	Four à hybridation Agilent (Agilent Technologies)	
Lavage	Universal microarray hybridation kit (Promega, France)	Succession de solutions de concentration décroissante en SSC et SDS	Gene Expression Wash Buffer Kit (Agilent Technologies)	

* CRB-GADIE : Centre de Ressource Biologique Génomique des Animaux Domestiques et d' Intérêt Economique

- ✓ Les lames Operon sont des lames utilisées par Ollier & *al.* (2007). Ces lames contiennent le set d'oligonucléotides bovins commercialisé par la société Opéron (n=8 329) auquel ont été ajoutés 50 oligonucléotides (appelé set « Clermont » dans le Tableau I.5, numéro d'accession GEO : GPL4594). Les oligonucléotides sont des 70mer, c'est à dire qu'ils comportent 70 nucléotides. Ces oligonucléotides ont été déposés (spottés) en dupliqua sur des lames UltraGAPS (Corning B.V. Life Sciences) sur la Plateforme « Biopuces » de la génopôle de Toulouse Midi Pyrénées INSA/DGBA. Ces lames comportent 19 200 spots correspondant à 8 379 séquences oligonucléotidiques différentes dont 6 656 (soit 79,4 %) sont des gènes différents ayant une annotation reconnue par le logiciel Ingenuity Pathway Analysis (IPA).
- ✓ Les lames du CRB (Centre de Ressources Biologiques GADIE[1], Inra, Jouy-en-Josas) ont été produites en associant le set d'oligonucléotides (70mer) bovins commercialisé par la société Operon (n=8 329) et le set d'oligonucléotides synthétisés par la société Illumina (n=13 257, numéro GEO : GPL2853). Ces oligonucléotides ont été spottés par le CRB-GADIE, sur des lames UltraGAPS (Corning B.V. Life Sciences). Les lames comportent 23 232 spots correspondant à 21 586 sondes (séquences) différentes dont 9 973 (soit 46,2 %) sont des gènes différents ayant une annotation acceptée par IPA. Le faible nombre de gènes par rapport au nombre de sondes s'explique en partie par le fait qu'il existe une certaine redondance entre les 2 sets commerciaux, et donc que certains gènes sont représentés par 2 (parfois davantage) sondes, mais aussi et surtout en raison de la pauvreté de l'annotation du génome bovin en séquences exprimées.
- ✓ Les lames Agilent sont fabriquées par la société Agilent (Agilent Technologies). La synthèse des oligonucléotides (60 mers) est réalisée *in situ*, c'est à dire directement sur la lame, par ajout successif de nucléotides, selon un principe dérivé du système mis au point par Hewlett Packard (HP) pour ses imprimantes à jet d'encre. Cette technologie permet de densifier le nombre de spots présents sur la lame de verre. Les lames Agilent comportent quatre réseaux de 44 407 spots correspondant chacun à 21 475 sondes différentes (synthétisées en dupliqua) dont 13 746 (64 %) sont identifiées comme correspondant à des gènes différents ayant une annotation reconnue par IPA.

[1] Génomique des Animaux Domestiques et d' Intérêt Economique

I.4.3. Les différentes étapes mises en œuvre pour la réalisation d'une analyse transcriptomique différentielle sur puces à ADN

L'objectif de cette partie est de présenter succinctement les étapes successives d'une expérimentation de puce à ADN.

I.4.3.1. La planification expérimentale

L'expérimentation sur puces à ADN comporte de nombreuses étapes qui sont autant de source de variabilités. Malgré l'importance de ces variabilités, certains biais peuvent être contrôlés et d'autres ne le peuvent pas (ils sont alors dits aléatoires). Les biais contrôlables peuvent être réduits en utilisant des procédures expérimentales strictes comme par exemple un même manipulateur pour une même expérience. Les biais aléatoires constituent un bruit qui réduit la puissance des analyses statistiques et la fiabilité des résultats.

La première étape de l'expérimentation sur puces à ADN pour un biologiste est donc de construire un plan expérimental approprié qui permettra de contrôler et corriger les biais liés à la technique et d'optimiser l'analyse statistique employée pour répondre à la question biologique posée (Mary-Huard *et al.*, 2006). L'objectif d'un plan expérimental est d'organiser l'analyse biologique pour en retirer l'information la plus précise dans un nombre limité d'expériences. Le plan expérimental est une construction de l'expérience qui répond aux exigences scientifiques et statistiques en tenant compte des contraintes de quantité, de temps et de budget. La puissance d'un plan expérimental peut être optimisée en minimisant la variance des estimateurs de l'effet étudié.

➢ La notion de répétitions

La sous-unité d'un plan expérimental est le nombre de répétitions effectuées pour chaque condition expérimentale comparée. Ce nombre est à limiter au minimum face aux contraintes matérielles. La notion de répétition comprend le nombre d'échantillons par individu et le nombre d'individus par condition expérimentale. Les répétitions correspondant aux échantillons sont appelées répétitions techniques, celles associées aux individus sont appelées répétitions biologiques. Une différence significative entre deux conditions expérimentales peut être détectée avec des répétitions techniques mais pas avec des répétitions biologiques, car la variabilité biologique est souvent plus élevée que la variabilité technique (Mary-Huard *et al.*, 2006). Cependant, les conclusions biologiques sont

complètement dépendantes du type de réplication utilisé. Généralement, l'objectif d'une expérimentation est de conclure quant à une population. Dans ce but, seules les répétitions biologiques sont valides puisqu'elle prennent en compte les variabilités individuelles. Les répétitions techniques sont alors utiles pour évaluer et corriger les biais techniques (Mary-Huard *et al.*, 2006). Dans le cas des puces à ADN « deux couleurs », les données sont appariées puisque deux conditions sont hybridées sur une même lame.

Dans la technologie des puces à ADN deux couleurs, on distingue deux modes de répétitions : le dye-swap et le dye-switch. Le dye-swap est une répétition technique d'une comparaison entre deux conditions expérimentales avec inversion de marquage (cf. Figure I.25). Sur une première lame, les cibles représentant la condition A sont marquées avec la cyanine 3 et les cibles de la condition B sont marquées avec la cyanine 5. Sur la deuxième lame, il y a inversion des fluorochromes c'est à dire que les cibles de la condition A sont marquées avec la cyanine 5 et que les cibles de la condition B sont marquées avec la cyanine 3. Dans un dye-swap, les individus hybridés sur la deuxième lame sont les mêmes que ceux utilisés sur la première lame. Le dye-switch est une variante du dye-swap. Dans le dye-switch, les individus utilisés sur la deuxième lame ne sont pas les mêmes que ceux utilisés sur la première lame.

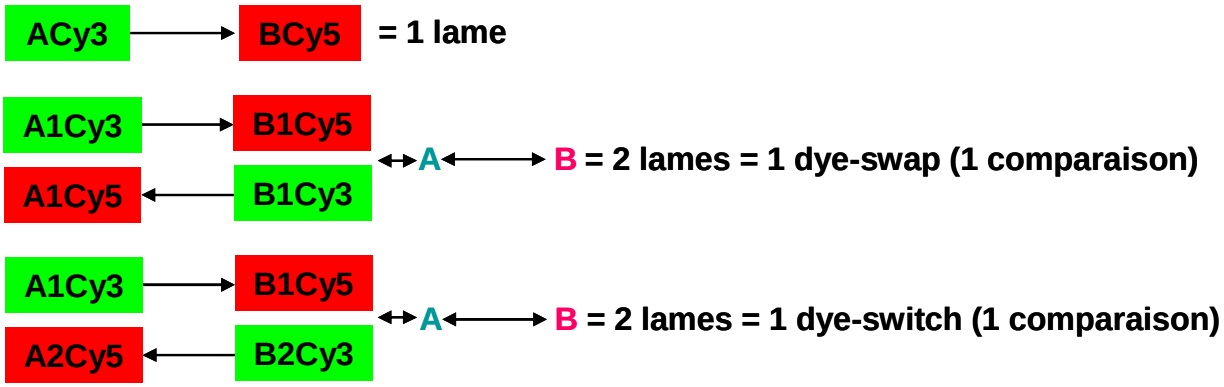


Figure I-25 : Nomenclature utilisée dans la schématisation des plans expérimentaux en technologie de puces à ADN « deux couleurs ».
Une flèche simple correspond à une lame sur laquelle sont hybridés les deux échantillons reliés par la flèche, chaque double flèche correspond à un dye-swap soit deux lames sur lesquels sont hybridés les deux échantillons reliés par la flèche. Chaque lettre correspond à une condition expérimentale, Cy : cyanine.

Le principe du dye-swap permet de dissocier l'effet du fluorochrome de l'effet lié à l'individu, puisque la cible complexe issue d'un individu est obligatoirement marquée avec les deux fluorochromes. Le principe du dye-switch ne permet pas de dissocier l'effet individu du biais lié au marqueur, car un individu est associé à un fluorochrome (Mary-Huard et al., 2006). Dans le cas où des dye-switch sont utilisés dans un plan expérimental comparant plus

de deux conditions, il est conseillé de marquer tous les individus avec les deux fluorochromes puis d'équilibrer le plan expérimental général en considérant deux échantillons (cyanine 3 et cyanine 5) par individu. Les études présentées dans ce travail de thèse, ont exclusivement fait appel à des expériences en dye-swap.

➢ Les différents types de plans expérimentaux

A chaque question biologique correspond un plan expérimental optimisé. Le plan expérimental doit concilier la question biologique avec la variabilité du matériel biologique disponible et le nombre de puces à ADN pouvant être utilisées.

Quelques types de plans sont présentés à la Figure I.26.

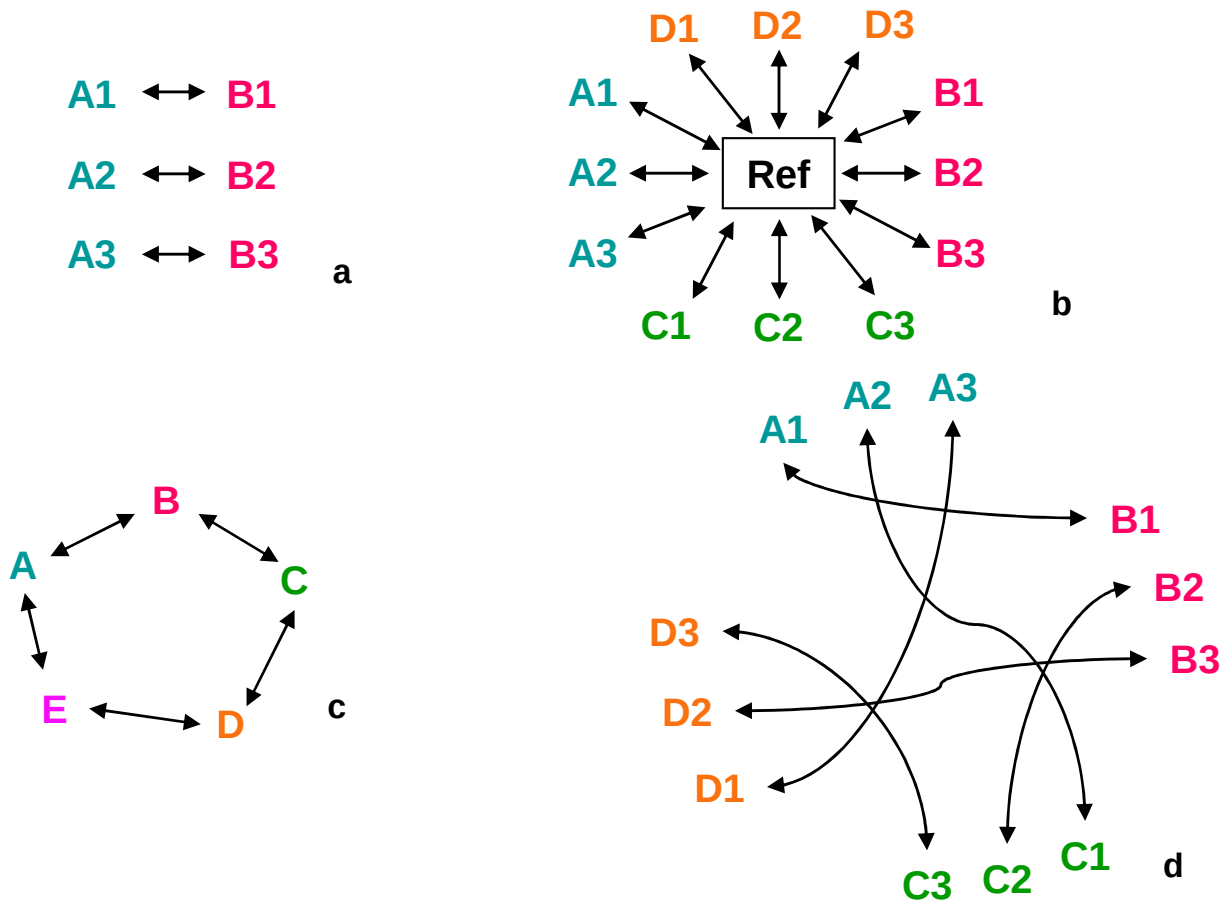


Figure I-26 : Représentation schématique de différents types de plans expérimentaux (adapté de Kerr & Churchill, 2001).

a) comparaison directe de deux conditions expérimentales A et B en dye-swap assortie de répétitions biologiques, b) comparaison indirecte de quatre conditions expérimentales (A, B, C et D) en référence à un référentiel commun, c) comparaison de cinq conditions expérimentales (A, B, C, D et E) à l'aide d'un plan expérimental en boucle, d) comparaison de quatre conditions expérimentales (A, B, C et D) à l'aide d'un plan expérimental en boucle avec renforcement des comparaisons deux à deux. Chaque double flèche correspond à un dye-swap soit à deux lames sur lesquelles sont hybridés les deux échantillons reliés par la flèche, chaque lettre correspond à une condition expérimentale, chaque chiffre correspond à une répétition biologique d'une condition expérimentale, Ref correspond à l'échantillon de référence.

Plan expérimental pour la comparaison de deux conditions

Le plan (a) permet la comparaison directe de deux conditions. Chaque comparaison entre deux individus de deux conditions expérimentales représente une réplication (ou répliqua) biologique. Une comparaison en dye-swap constitue une réplication technique. Ce schéma est conseillé dans le cas où les individus d'une même condition expérimentale sont assez différents entre eux : il faut alors augmenter le nombre de comparaisons pour dégager l'effet lié aux conditions expérimentales.

Plans expérimentaux pour la comparaison de plusieurs conditions expérimentales

Le plan (b) est un plan expérimental en référence qui compare indirectement plusieurs conditions expérimentales entre elles. Ce plan peut également être utilisé pour comparer deux conditions. L'avantage de ce plan expérimental est qu'il permet de rajouter des conditions en cours d'expérience. L'inconvénient de ce plan est qu'il y a une perte de puissance dans la comparaison des conditions entre elles, puisqu'il n'y a jamais comparaison directe entre deux conditions sur un même support. De plus, si ce plan est réalisé avec des dye-swap à chaque comparaison, il nécessite un grand nombre de lames. Le nombre de lames nécessaire peut être réduit en associant la référence à un fluorochrome et en marquant toutes les autres conditions avec l'autre fluorochrome, bien que ce processus ne permette pas de dissocier l'effet condition du biais lié au marquage.

Le plan (c) est un plan expérimental en boucle comparant plusieurs conditions simultanément. Il est très utilisé dans le cas d'études cinétiques et dans le cas où les individus ne sont pas trop différents entre eux (culture cellulaire), lorsqu'il n'y a pas de répétition biologique d'une même condition.

Le plan (d) enfin, est un plan expérimental en boucle comparant plusieurs conditions avec des répétitions biologiques pour chacune d'elles. Dans la Figure I.26, le plan est parfaitement équilibré, car chaque condition est comparée directement à toutes les autres. La variabilité liée à l'individu est corrigée par la boucle. Quand cela n'est pas possible (principalement lorsqu'il n'y a pas suffisamment de réseaux), il est possible de retirer certaines comparaisons directes entre deux conditions, là où elles sont les moins importantes. La comparaison entre ces deux conditions se fera alors indirectement à travers les comparaisons directes entre les autres conditions.

I.4.3.2. Préparation des cibles complexes, hybridation et lavage

La préparation des cibles complexes consiste à transformer les ARN totaux, extraits des tissus à comparer, en cible complexe qui puisse se lier de façon spécifique aux sondes présentes sur les puces à ADN. Cette phase du processus contient une étape de marquage qui dépend du type de lame utilisé et permettra de révéler l'existence d'un éventuel différentiel d'expression pour une séquence connue (sonde).

Il existe 3 types de marquage se distinguant par la position de l'étape de marquage par rapport à la rétro-transcription et par la présence ou non d'une étape d'amplification. On parle de marquage direct lorsque celui-ci se fait pendant la rétro-transcription et de marquage indirect lorsque celui-ci se fait après l'étape de rétro-transcription. Trois types de marquage ont été utilisés au cours de ce travail de thèse, en fonction des diverses lames utilisées (cf. Figure I.27).

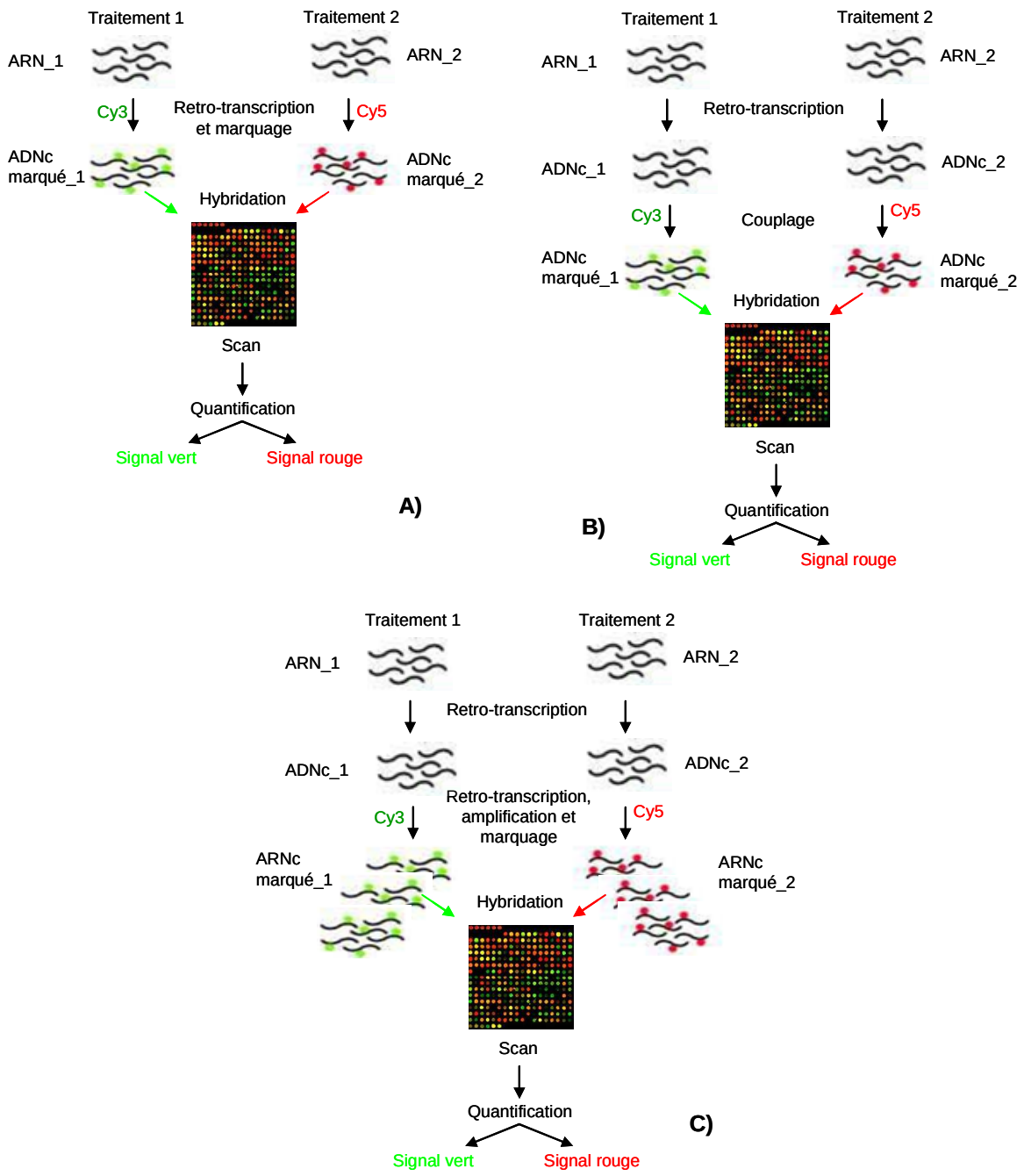


Figure I-27 : Les différents types de marquages en technologie de puces à ADN « deux couleurs ».
A) marquage direct, B) marquage indirect, C) marquage associé à une amplification. ADNc : Acide Désoxyribonucléique complémentaire, ARN : Acide Ribonucléique, ARNc : ARN complémentaire, Cy3 : cyanine 3, Cy5 : cyanine 5.

Le marquage direct consiste à rétro-transcrire l'ARN total en ADN complémentaire (ADNc) en incorporant directement dans la séquence nucléotidique néosynthétisée des dCTP (deoxytidine triphosphate) marqués soit par la cyanine 3, soit par la cyanine 5.

Le marquage indirect consiste à rétro-transcrire dans une première étape les ARN totaux en ADNc, en incorporant une catégorie de nucléotides portant un groupement amino-allyl. Dans une deuxième étape, les ADNc sont marqués par les cyanines 5 et les cyanines 3 qui sont

couplés à la cible via les groupements amino-allyl. Le marquage indirect présente l'avantage de pouvoir conserver des ADNc non marqués.
Ces deux types de marquage utilisent entre 5 à 20 µg d'ARN totaux.

Le marquage associé à une étape d'amplification consiste à rétro-transcrire dans une première étape les ARN totaux en ADNc puis à synthétiser des ARN complémentaires (ARNc) marqués. La synthèse des ARNc procède par transcription des ADNc, tout en incorporant directement des dCTP porteurs de fluorochromes. L'étape de transcription *in vitro* correspond à l'étape d'amplification. L'avantage de cette technique est qu'elle est peu consommatrice d'ARN totaux. En effet, à partir de 500 ng d'ARN totaux il est possible de produire suffisamment de cibles complexes marquées pour hybrider plusieurs lames.
La qualité d'un marquage s'évalue en mesurant les longueurs d'onde à 260, 550 et 650 nm et en calculant les fréquences d'incorporation (FOI) des cyanines selon les modes de calcul proposés par le fournisseur dans le protocole de marquage. Les seuils d'acceptabilité des échantillons sont également recommandés par le fournisseur.

Deux types de transcriptions inverses ont été utilisés dans nos études. La première utilisait des hexamères et permettant la rétro-transcription aléatoire de tous les ARN. La deuxième utilisait des oligo(dT) et permettait la rétro-transcription des ARNm uniquement (c'est à dire ceux contenant une queue poly(A)) en commençant par l'extrémité 3'.

L'étape d'hybridation est celle qui met en contact les cibles complexes marquées avec les sondes présentes sur les lames.

Selon le type de lame, un prétraitement avant hybridation peut être nécessaire. Cette étape est appelée pré-hybridation et sert à bloquer les sites libres de captation d'ADN sur les lames afin de limiter le bruit de fond.

Les protocoles d'hybridation des cibles complexes (tampon, température…) et de lavage des lames après hybridation sont à adapter au type de support et la nature des sondes.

I.4.3.3. Acquisition des images

Une fois l'hybridation réalisée, l'acquisition des images des puces à ADN se fait en scannant les lames sur un imageur approprié. Le principe des scanners est d'exciter les fluorochromes fixés sur les sondes par un laser rouge (à 635 nm, excitant la cyanine 5) et un laser vert (à 532 nm, excitant la cyanine 3), d'amplifier la fluorescence émise par les cyanines excitées à l'aide d'un tube photomultiplicateur (PMT) et de mesurer puis de convertir en

signal électrique et en données chiffrées cette fluorescence sur un convertisseur analogique digital. Certains paramètres du scanner doivent être choisis ou optimisés en fonction de la lame scannée tels que la résolution du scan (2, 3, 5 ou 10 µm) dépendant de la densité des oligonucléotides, la puissance des lasers et la zone de lecture.

I.4.3.4. Quantification des signaux

La quantification du signal rouge et vert de chaque spot s'opère grâce à des logiciels de traitement d'image dédiés. La première étape consiste à poser une grille spécifique de chaque lame sur les images permettant de localiser et de donner une identification à chaque spot. La deuxième étape consiste à sélectionner les options de quantification. Ces options concernent les critères d'élimination des spots : pour non uniformité du spot ou du bruit de fond, non-existence d'un spot (flags), ou de résolution des spots en terme de pixels et enfin de mode de calcul du signal (moyenne ou médiane des pixels du spot et du bruit de fond). La troisième étape consiste à exécuter la quantification, ce que réalise automatiquement les logiciels « Imagene » (Biodiscovery, USA) et « Feature Extraction » (Agilent technologies).

I.4.4. Analyse des données issues des puces à ADN

I.4.4.1. Normalisation des données

Avant la normalisation des données, il est possible d'éliminer certains spots non conformes des fichiers de quantification. Les critères d'élimination sont fixés pendant cette étape (cf. paragraphe I.4.3.4.). Afin de conserver toute l'information contenue sur les lames, dans le cadre de nos études, aucun critère d'élimination n'a été pris en compte. Une vérification sur les gènes utilisés pour l'interprétation biologique des données a été effectuée *a posteriori*, après analyse statistique.

La normalisation a pour objectif de corriger les biais systématiques des données et de supprimer, quand cela est possible, toute influence non biologique sur ces données (Bogard *et al.*, 2008). La normalisation oblige souvent à transformer l'ensemble des données brutes en données logarithmiques (généralement en base 2) car l'amplitude de variation des intensités brutes est très large. En effet, certains spots fournissent un signal très faible alors que d'autres peuvent conduire à un signal saturant. La détection des biais non biologiques se fait en comparant graphiquement deux entités calculées : le ratio des intensités, $M = \log_2(R/V)$ et la moyenne des intensités, $A = ½\log_2(R/V)$ ou R représente l'intensité du signal médian dans le canal rouge et V l'intensité du signal médian dans le canal vert.

Deux types de correction dans l'étape de normalisation peuvent être appliquées. La première correction concerne tous les types de lame et consiste à corriger les signaux de l'effet lié à la différence de comportement des fluorochromes. Cette normalisation, appelée Lowess (ou Loess), suppose que peu de gènes s'expriment différemment entre les deux conditions expérimentales comparées sur la lame. Sur les graphes M-A (M-A plot) des données brutes (Figure I.28), le nuage de points en forme de « banane » indique que les données ne sont pas symétriques, car non centrées sur zéro. Ceci est le reflet de courbures liées à une non linéarité des intensités. Le canal rouge (Cy5) a un signal plus fort aux basses intensités et le canal vert (Cy3) présente un signal plus fort aux fortes intensités. La méthode Lowess corrige ces courbures et fournit un nuage de points centrés sur une droite (Figure I.28B) (Bogard *et al.*, 2008; Mary-Huard *et al.*, 2006). La correction de ce biais se fait localement pour chaque spot en appliquant un ajustement « Lowess » à l'intensité du signal médian du spot. Le signal de chaque point est ajusté au signal des points qui l'entourent. L'option choisie lors de l'application d'une normalisation « Lowess » est le paramètre *f* qui détermine la fraction des données à utiliser pour lisser chaque point (Yang *et al.*, 2002a). La deuxième étape de normalisation se fait uniquement sur des lames spottées (type Operon ou CRB). Sur les lames spottées, il y a présence de blocs représentatifs des aiguilles d'un spotter. D'une aiguille à l'autre, il existe des différences d'homogénéité des spots (forme, concentration,…), il est donc nécessaire de normaliser les spots d'un même bloc pour pouvoir comparer deux gènes situés sur deux blocs différents. L'effet « bloc » se visualise sur un box-plot (cf. Figure I.28C). La normalisation par bloc consiste à retirer pour chaque signal de chaque spot le signal médian du bloc dans lequel ce spot se trouve (Yang *et al.*, 2002a). Après la normalisation par bloc, tous les spots d'une lame sont comparables entre eux (cf. Figure I.28D). Cette normalisation n'est pas nécessaire pour les lames de type Agilent (synthèse des sondes oligonucléotidiques *in situ*).

Dans le cadre de cette thèse, ces deux étapes de normalisation ont été réalisées à partir de la fonction *normalisation* du package « anapuce », développé sous le logiciel R par Julie Aubert (UMR 518, AgroParisTech). Nous avons dû procéder à une adaptation de cette fonction pour permettre la lecture des fichiers issus du logiciel « Feature Extraction ».

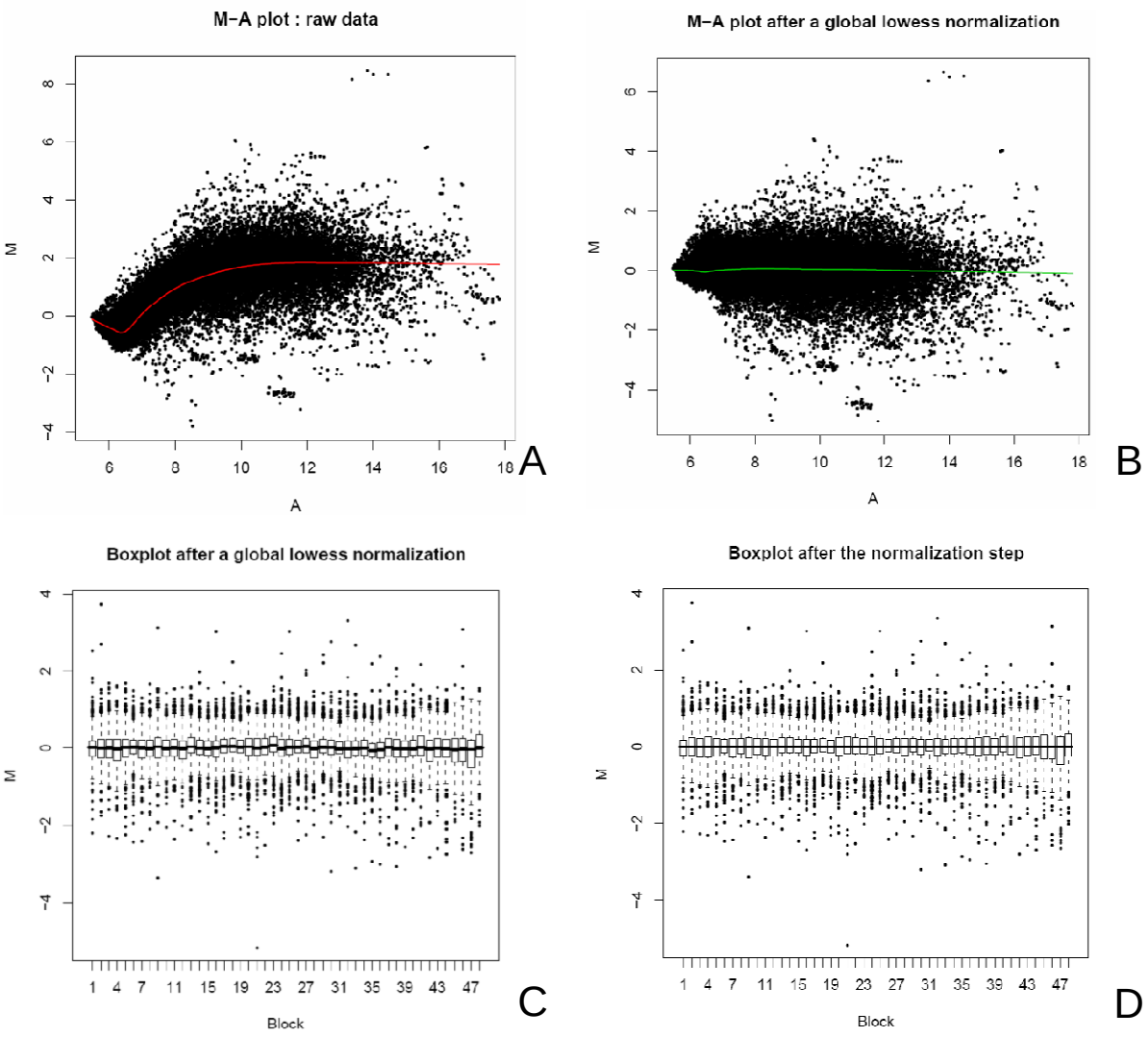


Figure I-28 : Exemples de graphes réalisés dans la normalisation de données issues de la technologie de puces à ADN « deux couleurs ».
(A) M-A plot des données brutes d'une lame de type Agilent. (B) M-A plot des données après normalisation « Lowess » d'une lame de type Agilent. (C) Boxplot des données après normalisation « Lowess » d'une lame de type CRB. (D) Boxplot des données après normalisation « Lowess » et par bloc d'une lame de type CRB.

Avant l'analyse statistique du jeu de données, il est nécessaire de vérifier la cohérence générale des données normalisées venant de plusieurs lames d'une même expérience. Cette vérification peut se faire (i) en calculant les coefficients de corrélation entre les lames, permettant de vérifier que les swap sont bien opposés (coefficient de corrélation négatif entre les deux lames d'un dye-swap) ; (ii) en calculant les coefficients de corrélation entre les intensités des signaux normalisées de toutes les lames, permettant de vérifier que l'effet lame n'est pas supérieure à l'effet échantillon et que les répétitions biologiques d'une même condition sont bien corrélées ; (iii) en traçant des boxplots permettant de visualiser rapidement

si les lames sont comparables entre elles. Ces vérifications permettent d'éliminer certaines lames incohérentes d'une étude.

I.4.4.2. Analyse statistique d'un jeu de données

Les informations données ci-après s'inspirent de la revue de Mary-Huard et al., 2006 construite, notamment à partir des références suivantes : Butte, 2002 ; Leung & Cavalieri, 2003 ; Nadon & Shoemaker, 2002 ; Quackenbush, 2001.

L'objectif d'une analyse différentielle est d'identifier les gènes différemment exprimés ou les gènes dont le niveau d'expression diffère d'une condition à l'autre. Ce type d'analyse comporte trois étapes : (i) la définition de la statistique qui évalue la différence d'expression entre deux conditions, (ii) la définition d'une règle de décision basée sur cette évaluation pour déclarer si le gène est différemment exprimé ou pas, (iii) le contrôle de la probabilité de la prise de la mauvaise décision (risque d'erreur). La question statistique posée est souvent « est-ce que l'expression d'un gène donné diffère entre une condition A et une condition B ? ». L'hypothèse nulle H_0 est alors l'absence de différence entre les deux conditions ($\mu_D = \mu_A - \mu_B = 0$, c'est l'égalité des variances estimées). Cette hypothèse est testée avec un t-test (ou test de Student) en considérant α, l'erreur de type I, c'est à dire la probabilité de déclarer un gène différemment exprimé alors qu'il ne l'est pas (faux positif) et β l'erreur de type II, la probabilité de déclarer un gène non différemment exprimé alors qu'il l'est. L'objectif de la règle de décision est de faciliter le choix mais aussi de contrôler la probabilité de se tromper.

La valeur de la statistique d'un test de Student dépend de deux quantités : la différence moyenne des valeurs d'expression entre les deux conditions et la variabilité de cette différence d'expression (Mary-Huard *et al.*, 2006). Il est donc nécessaire de modéliser la variance pour chaque gène. Il existe trois modélisations de la variance. La première considère une variance spécifique par gène, c'est à dire que chaque gène présente une variabilité spécifique d'expression entre les deux conditions. Cette modélisation est proche de la réalité d'un point de vue biologique, mais difficile à obtenir d'un point de vue statistique. Tusher & *al.* (2001) ont proposé un modèle d'analyse de ce type appelé Significant Analysis of Microarray (ou SAM) qui ajoute une variance constante à la variance estimée afin de palier au faible nombre de répliquas techniques et donc à une erreur de l'estimation de la variance. La deuxième modélisation considère une variance commune à tous les gènes. Cette modélisation est avantageuse pour l'analyse statistique mais est trop simpliste et peu réaliste d'un point de vue biologique. Dans ce cas, la variance représente la moyenne de l'expression de tous les gènes et est fixée pour tous les tests de Student. La dernière modélisation est intermédiaire,

elle considère que les gènes peuvent être placés dans des groupes de même variance (Delmar *et al.*, 2005). C'est un modèle de variance mixte où chaque gène a la variance du groupe de gènes dans lequel il se trouve.

Le problème majeur de l'analyse des données de lame de verre est la répétition du test de Student autant de fois qu'il y a de gènes sur la lame de façon indépendante. Le nombre de gènes déclarés, par erreur, différemment exprimés dans les deux tissus est alors proportionnel au seuil α retenu pour le test, ce qui peut engendrer un nombre important d'erreurs. Par exemple, si la lame contient 10 000 gènes et que l'erreur est de 5%, 500 gènes pourraient être déclarés différemment exprimés alors qu'ils ne le sont pas. Pour éviter cela, il faut considérer un test unique appelé test multiple, dont la procédure est associée à une hypothèse nulle complète H_0^C « il n'y a pas de différence d'expression entre les conditions A et B pour aucun des gènes ». Le test multiple contrôle alors le risque global associé à une hypothèse nulle complète (H_0^C). Des corrections des probabilités associées au test global sont alors possibles. La première, appelée « Family Wise Error Rate » (FWER), est la procédure de Sidak et Bonferroni, qui considère que dans la liste des gènes déclarés différemment exprimés par le test multiple, il y a une probabilité α (erreur de type I) d'obtenir au moins un gène faux-positif. La seconde, appelée « False Discovery Rate » (FDR), est la procédure de (Benjamini & Hochberg, 1995), qui considère que dans la liste de gènes déclarés différemment exprimés par le test multiple, il y a α (probabilité de l'erreur de type I) gènes faux-positifs.

Pour nos études, nous avons utilisé la fonction *DiffAnalysis* du package « anapuce » qui fournit un fichier par type de modélisation de la variance contenant les deux méthodes de correction des probabilités. Lorsque certaines séquences étaient dupliquées sur la lame (plusieurs spots correspondant à une même sonde), la fonction *DiffAnalysis* effectuait la moyenne de leur intensité.

I.4.4.3. Clustering

Le clustering est une manière simple d'ordonner le nombre important de données issues des puces à oligonucléotides. Il permet de regrouper des gènes et/ou des échantillons en fonction de leur profil d'expression. Ce regroupement peut s'effectuer avant analyse statistique pour visualiser l'ensemble des données ou après analyse statistique pour ne travailler que sur la liste des gènes dont l'expression diffère d'une condition à l'autre. Le choix de l'algorithme de regroupement a un impact sur le résultat final ; ce n'est donc pas une représentation objective de l'information. Il y deux méthodes de clustering, celle basée sur la distance (« distance-based ») et celle basée sur le modèle (« model-based ») (Mary-Huard *et*

al., 2006). Cette partie ne décrira que la méthode basée sur la distance car c'est la seule méthode que nous ayons utilisée. Comparativement à cette méthode, celle basée sur le modèle est plus performante, plus informative et plus reproductible mais nécessite de maîtriser les algorithmes de calculs. Un clustering est considéré satisfaisant lorsque les groupes sont homogènes et bien séparés, c'est à dire que la variance intragroupe doit être comparable et qu'il doit y avoir une grande variabilité intergroupe. Les deux types de méthodes de clustering basées sur la distance sont le clustering hiérarchique, qui est une approche visant à minimiser la variabilité intragroupe et le K-means clustering, qui construit des clusters homogènes en caractérisant chaque cluster par un point central (moyenne), en affectant chaque gène au cluster le plus proche et en recalculant la moyenne de ce cluster. Lors de l'utilisation de clustering basé sur la distance, il est conseillé d'essayer plusieurs regroupements en changeant les paramètres (nombre de clusters, groupes de gènes au départ…), et de ne retenir que le regroupement le plus informatif.

I.4.4.4. Interprétation biologique des données

Une fois établie la liste des gènes déclarés différemment exprimés ou la liste des groupes de gènes de même profil d'expression, l'étape suivante est d'interpréter biologiquement ces données.

Chaque séquence oligonucléotidique de la lame est associée à un gène ou à un EST. Les avancées en matière de séquençage des génomes, notamment du génome bovin, et dans la réannotation des gènes, surtout en terme de fonction biologique, obligent souvent à réannoter les sondes présentes sur les lames avant de réaliser l'interprétation biologique. S'agissant des lames utilisées au cours de nos travaux, les réannotations sont faites à partir des outils développés sous Biomart par SIGENAE (Système d'Information des GENomes des Animaux d'Elevage). A chaque séquence oligonucléotidique de la lame correspondait un numéro de contig « Sigenae » qui permettait, lorsque cela était possible, de remonter à une annotation « GeneBank » bovine ou « Unigene » humaine. Avec ces annotations, il était alors possible d'utiliser les bases de données d'interprétation biologique. Lorsqu'il était impossible d'obtenir une annotation dans Biomart à partir des numéros de contig « Sigenae », l'annotation était alors faite manuellement en s'appuyant sur les informations fournies avec les lames et sur la base de données du NCBI (National Center for Biotechnology Information).

Dans ces travaux de thèse, deux bases d'interprétation ont été utilisées : la base de données DAVID du National Institute of Allergy and Infectious Diseases (NIAID), NIH (Dennis *et al.*, 2003) et le logiciel Ingenuity Pathway Analysis (IPA). DAVID est une base de

données dont l'accès est libre, pouvant traiter des annotations bovines. Cette base de données ne présente pas de lieu de sauvegarde et de stockage des données chargées. A chaque analyse il est nécessaire de recharger les fichiers. IPA est un logiciel coûteux qui utilise des annotations humaines, rat ou souris. Ce logiciel permet la conservation des données pendant toute la durée de la licence. Comme IPA possède de l'information bibliographique sur plus de 14 ans, il ne restreint pas l'interprétation biologique aux annotations fonctionnelles de Gene Ontology (GO). IPA propose également de nombreuses options d'analyse telles que la possibilité de fixer des limites d'analyse, de créer des réseaux, d'interroger les fichiers de données sur une liste de gènes connue…

Les informations fournies par une base de données d'interprétation biologique ne sont pas absolues. Il est conseillé de vérifier dans la littérature que les gènes présents dans la liste étudiée sont bien associés à la fonction biologique proposée par la base de données.

II. Chapitre 2 : étude transcriptomique de la différenciation terminale du tissu mammaire caprin

Les facteurs et les mécanismes locaux participant à la différenciation terminale de la glande mammaire des ruminants au cours de la gestation sont relativement mal connus. Toutefois, dans la littérature, il existe quelques études effectuées chez les ruminants et portant sur la transition gestation-lactation (Baldwin, 1966; Finucane *et al.*, 2008; Mellenberger *et al.*, 1973) ou sur l'effet de différentes hormones telles que les hormones thyroïdiennes (Capuco *et al.*, 2008).

Devant l'absence de connaissances sur la différenciation des tissus mammaires des ruminants, notamment dans des stades précoces de la gestation, nous avons entrepris une analyse globale de la différenciation terminale des tissus mammaires caprins. Pour ce faire, nous avons choisi une approche de trancriptomique visant à tracer les profils d'expression de milliers de gènes sur cinq stades de développement, quatre en gestation et un en lactation.

L'analyse statistique de données issues d'une approche transcriptomique suppose que les conditions comparées directement soient globalement peu différentes. Or, dans notre expérience, nous souhaitions comparer cinq stades de développement de la glande mammaire dont un stade précoce de la gestation (46 jours, G46) avec un stade établi de la lactation (40 jours, L40). Pour ce faire, nous avons choisi une étude cinétique établie sur un plan expérimental en boucle qui comparait tous les stades entre eux. Pour pallier aux exigences imposées par l'analyse statistique, nous avons retiré la comparaison directe entre G46 et L40 du plan expérimental de l'étude globale. Cette comparaison a fait l'objet d'une autre expérience.

Ce chapitre a donc été scindé en 2 sous parties, correspondant aux deux analyses : la première trace les profils d'expression de plusieurs milliers de gènes sur cinq stades de développement de la glande mammaire ; la seconde compare directement un stade précoce de la gestation (G46) avec un stade de lactation établie (L40), essentiellement dans le but de caractériser, par des profils d'expression respectifs, une transition fonctionnelle marquée.

II.1. Projet de l'article 1 : La différenciation terminale des tissus mammaires caprins au cours de la gestation requière l'expression de gènes du système immunitaire

Résumé, de l'article 1 (soumis pour publication à Physiol. Genomics)

La différenciation terminale des tissus mammaires en un épithélium fonctionnel qui synthétise et sécrète du lait a lieu pendant la gestation. Les mécanismes moléculaires sous-jacents à ce processus complexe restent partiellement incompris, surtout chez les ruminants. Une analyse transcriptomique de cinétique utilisant la technique des lames de verre pour comparer 5 stades de développement physiologiques (gestation et lactation) a permis d'étudier le processus général et final de différenciation de la glande mammaire des ruminants. Afin de suivre les profils d'expression génique au cours de cette différenciation, un plan expérimental en boucle a été conçu et validé. La comparaison des stades physiologiques s'est faite sur 9 dye-swaps en utilisant des lames 22K bovine (CRB-GADIE, INRA, France) et 3 individus nullipares (gestation) ou primipares (lactation) par stade. L'analyse statistique a mis en évidence que 1 696 gènes ont varié significativement au moins une fois dans cette étude. Ces gènes ont été répartis dans 19 groupes selon leur profil d'expression. L'identification des fonctions biologiques avec le logiciel Ingenuity Pathway Analysis (IPA) indiquait plusieurs similarités avec les stades physiologiques bien décris chez la souris. Ainsi par exemple on a pu identifier l'expression à mi-gestation des gènes spécifiant quelques lactoprotéines (caséines κ, β et α_{s2}) ; l'augmentation, avec la lactation, de l'expression des gènes impliqués dans le métabolisme et la sécrétion des lipides (*BTN1A1*, *GlyCAM-1* et *SCD*). Pendant la première moitié du temps de gestation, les tissus mammaires caprins sont des tissus croissants auto-génératifs, dans lesquels les processus de remodelage et de différenciation sont activés. La deuxième partie de la gestation est principalement consacrée à la prolifération. Un phénomène original, concerne dès la mi-gestation, l'expression, dans les tissus mammaires caprins, des gènes impliqués dans le système immunitaire, accréditant l'idée émergente d'une origine immunitaire de la glande mammaire. Ces données suggèrent que l'induction de la différenciation a lieu dans un stade très précoce de la gestation. Ce processus est toutefois crucial pour l'obtention d'une glande mammaire saine et productive.

Terminal differentiation of goat mammary gland during pregnancy requires the expression of immune functions genes

Authors:

FAUCON F.[1,2], REBOURS E.[1], BEVILACQUA C.[1], HELBLING J.-C.[1], AUBERT J.[3], MAKHZAMI S.[1], DHORNE-POLLET, S.[1], ROBIN S. [3] and MARTIN P.[1]*

Affiliation:

[1] INRA, UMR1313, Génétique animale et Biologie intégrative, équipe LGS, F-78352 Jouy-en-Josas

[2] Institut de l'Elevage, F-75012 Paris

[3] UMR INAPG/ENGREF/INRA 518, Mathématiques et Informatique Appliquées, F-75005 Paris

II.1.1.Introduction

Morphological development of the mammary gland starts during the fetal life of each female and is intermittent during the life with isometric growth phases during the post-natal period and after puberty, and allometric growth phases during puberty or at the onset of puberty for ruminants and during pregnancy (Hovey *et al.*, 2002). The terminal differentiation of the mammary tissue into an epithelium capable to synthesize and secrete milk occurs only during pregnancy and is under the control of multiple hormones and paracrine factors. Progesterone, the main placental hormone during pregnancy, could regulate the lobulo-alveolar setting, the alveolar proliferation and activates the synthesis of milk components already at mid-pregnancy in mice (Neville *et al.*, 2002) and bovine (Connor *et al.*, 2007) mammary tissues. Prolactin, a pituitary hormone, triggers lactation and is essential for its maintenance. In addition it plays a crucial role in the early stage of alveolar proliferation, in mice, before apparently being replaced by placental hormones (Neville *et al.*, 2002). The placental lactogen hormone is activated in late pregnancy and is involved in the goat mammary gland development by inducing alveolar epithelial cells differentiation and proliferation (Akers, 1985; Hennighausen & Robinson, 2001). The cortisol which is the main bovine glucocorticoid acts in the lobulo-alveolar differentiation at the parturition (Tucker, 2000). It was recently reported that IGF-1, a local factor secreted either by the stroma and/or the adipocyte, after activation by the Growth Hormone (GH), participates to epithelial cells proliferation in heifer mammary tissue (Connor *et al.*, 2007). Some other local factors, well described for mice, involved in the mammary gland and the alveolar development such as

RANKL, inhibin β, TGF β, cyclin D1, A-myb and C/EBP β are secreted by the epithelial cell as well as by the stroma (reviewed in Hennighausen & Robinson, 1998; Hennighausen & Robinson, 2001).

After puberty, rodent's mammary gland is composed of many ramified ducts finished by end buds (Terminal End Bud: TEB) and surrounded by a mammary fat pad (Hovey *et al.*, 1999; Hovey *et al.*, 2002). In rodents, the first third of pregnancy fits with a lobulo-alveolar growth, followed by a secretory differentiation in the second half of pregnancy. The last half of pregnancy is characterized by a cellular growth (Hennighausen & Robinson, 1998; Pitelka *et al.*, 1973). The secretory activation which is the ultimate step of the mammary gland differentiation occurs only at parturition and is characterized by the expression of genes involved in lipid and lactose biosynthesis (Anderson *et al.*, 2007; Rudolph *et al.*, 2003) eventhough there is a noticeable expression of caseins and/or α-lactalbumin (mRNA and proteins) in the mid-pregnant mammary tissue (Kanazawa & Kohmoto, 2002; Nakhasi & Qasba, 1979; Robinson *et al.*, 1995; Rudolph *et al.*, 2003).

At the beginning of pregnancy, ruminants mammary glands are made of an epithelium associated with a fibroblastic connective tissue with relatively few lipid-filled adipocytes (Hovey *et al.*, 1999). The mammary gland is structured by a cistern, a primary duct that ramifies in secondary and tertiary ducts. At the end, clusters of ductules form the so-called Terminal Ductule Lobular Units or TDLU (Akers, 1990; Hovey *et al.*, 1999; Hovey *et al.*, 2002). The differentiation of the mammary gland into a secretory tissue, during pregnancy, has not been described in details for ruminants. Only one morphological description of the bovine mammary tissue development and differentiation has been reported, 30 years ago by Swanson and Poffenbarger (1979). The authors showed that the bovine mammary tissue has already lobulo-alveolar structures visible at mid-pregnancy and that these structures contain some secretion material. The mammary parenchyma is developing in a continuous way in the mammary gland. At parturition, the mammary epithelium is occupying all the mammary space, the alveolar structures are fully developed and contain more defined secretion material (fat globules and casein micelles). The latest studies on ruminants mammary gland development and differentiation have more than 25 years (Anderson *et al.*, 1981). They show that the mammary gland weight, DNA and RNA levels increased during pregnancy, reflecting an increase in the cellular activity. Enzymatic activity increases in the bovine mammary tissue several weeks before the initiation of lactation (Baldwin, 1966). Moreover, enzyme activities for lactose and fatty acid synthesis increase between seven days pre-partum and seven days post-partum (Mellenberger *et al.*, 1973).

Whereas a lot of works have been done, during the last few years, on mouse mammary gland development including gene expression profiling experiments (Clarkson *et al.*, 2003; Master *et al.*, 2002; Naylor *et al.*, 2005; Rudolph *et al.*, 2003), by contrast there are very few descriptions of general mechanisms and biological changes underlying ruminants mammary gland terminal differentiation, during pregnancy. Microarray analysis on cow mammary tissue shows that sixteen genes are preferentially expressed during the development of the mammary gland (Suchyta *et al.*, 2003) and that the switch between the end of pregnancy and lactation is characterized by a down-regulation of genes encoding cell proliferation factors and an increase in the expression of genes involved in milk synthesis (Finucane *et al.*, 2008). In addition, to draw a network of genes that participate in coordinating milk fat synthesis and secretion, Bionaz and Loor (2008) have quantified the expression of 45 relevant genes from the late pre-partum/non lactating period through the end of subsequent lactation. Recently, expression patterns for genes that can alter mammary tissue sensitivity to galactopoietic thyroid hormones were evaluated during the last fifty days before expected parturition and during the first weeks of lactation (Capuco *et al.*, 2008). However there is still a lack of global expression analysis at different time points of the final differentiation process of the ruminant mammary tissues.

Here we report the identification of genes that are differentially expressed during terminal differentiation of the goat mammary tissue to yield a functional epithelium capable to synthesize and secrete milk. In addition, longitudinal mRNA expression analyses performed using oligoarrays for gene expression profiling during the course of pregnancy provide gene networks and clues to go further into the understanding of mechanisms by which the goat mammary tissue acquires its secretion phenotype.

II.1.2.Materials and methods

Animals and tissue sampling

Pregnant or primiparous lactating goats, either Saanen or Alpine from the Brouessy INRA experimental farm (France), were euthanized in safety and painless conditions, following procedures in compliance with the INRA Animal Care Committee guidelines. Mammary tissues were directly taken from the freshly isolated mammary gland. Approximately 1 cm^3 sections of tissue were immediately frozen in liquid nitrogen and kept at -80°C. Five different development stages of the mammary epithelium, thereafter called time points, corresponding to the following stages: 46, 70, 90 and 110 days of pregnancy and 40 days of lactation, were studied. The corresponding tissue samples were respectively identified as P46, P70, P90,

P110 and L40. Mammary glands from three different goats were used to characterize each pregnant or lactating stage.

RNA extraction

Frozen mammary tissue samples were grinded in liquid nitrogen, using a Freezer Mill 6750 (Fisher Scientific Bioblock). Total RNAs were extracted by mixing the resulting powder with Trizol reagent following manufacturer's instructions (Invitrogen, France). Resulting aqueous phase from each sample was purified on QIAGEN RNeasy mini-columns. After DNaseI treatment (Qiagen, France) RNAs were eluted and assayed by determination of 260/280 nm ratio using a nanodrop ND-1000 spectrophotometer (Nanodrop Technology, Nyxor Biotech, France). The quality of each RNA sample was evaluated using an Agilent 2100 Bioanalyzer with 6000 Nano LabChip kits (Agilent Technologies, France). The RIN factor, corresponding to the RNA Integrity Number (Schroeder *et al.*, 2006), was set to an inferior value of 7.5 for sample rejection.

RNA labelling

RNA samples were labelled using Superscript™ Indirected cDNA labelling protocol (Invitrogen, France) according to manufacturer's instructions. Briefly, 10 µg of total RNA were first reverse transcribed using SuperScript™ III Reverse Transcriptase enzyme and amino-modified dNTPs. The synthesized cDNAs were separated from the non-incorporated amino-modified dNTPs on S.N.A.P.™ columns. They were labelled with Cy3™ or Cy5™ dyes (Amersham Bioscience, GE Healthcare, France) and further separated from non-incorporated dyes on S.N.A.P.™ columns. The frequencies of dye incorporation (FOI, rate per thousand of labeled nucleotides) and the concentration of cDNA were estimated for both dyes by measuring the absorbance at 260 nm, 550 nm and 650 nm on the Nanodrop spectrophotometer. The following ratios were calculated FOI (Cy3) = (Abs550 X 324.5) / (Abs260 X 5.55), FOI (Cy5) = (Abs650 X 324.5) / (Abs260 X 9.25) and a superior limit of 20 was set to keep further any cDNA sample. Purity and quality of cyanine incorporation of each sample was finally checked by electrophoresis on a 1% agarose gel. The fluorescence was followed by Typhoon 9410 (GE Healthcare, France).

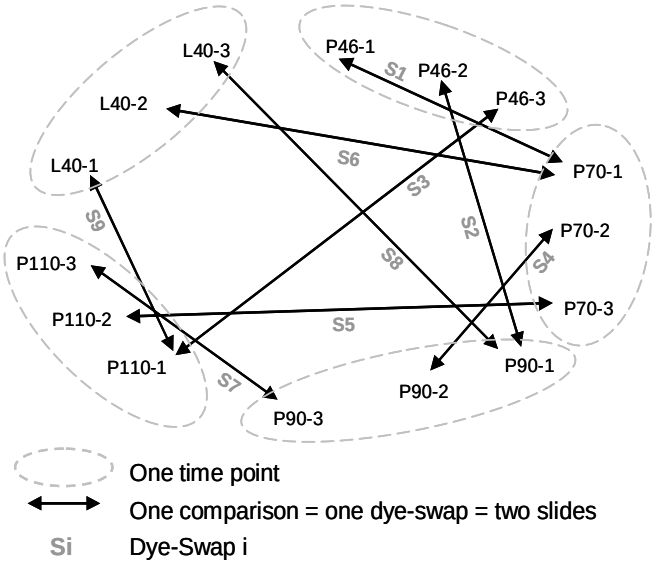
Microarray processing and data analysis

Since DNA microarrays were not available for the goat species and given the huge amount of work and money required to develop and produce a microarray for an "orphan" species we

decided to use an heterologous system, *i.e.* an oligoarray developed for a closely related species (bovine), for gene expression profiling. Such an approach has been already and successfully applied to study the impact of extreme genotypes at a casein locus (*CSN1S1*) on the goat mammary transcriptome (Ollier *et al.*, 2007). Microarrays used in this study were 22k bovine oligonucleotide probe arrays spotted by the *Centre de Ressources Biologiques - Génomique des Animaux Domestiques et d'Intérêt Economique* (CRB-GADIE, INRA, Jouy-en-Josas, France). Over the 23,232 probes spotted on the 22k bovine slide, 960 were empty spots, 260 were spots containing buffer, 8,709 spots were oligonucleotides from the Bovine Genome Oligoset V1.1 (Operon Biotechnologies) and 13,303 spots were oligonucleotides from a pre-existing set (NCBI GEO: GPL2853), developed starting from a bovine embryo cDNA library (Everts *et al.*, 2008). These two bovine oligonuclotide sets will be identified thereafter as “Operon” and “Illumina” sets, respectively.

Prior to hybridization, microarrays were pre-hybridized for 30 minutes at 50°C in a filtered solution of 1% BSA, 3.5X SSC, 0.1% SDS and twice rinsed in ultra pure water for 5 minutes under orbital agitation. Each microarray was co-hybridized with two fluorescently labeled (Cy3 and Cy5) cDNA samples. An evaluated amount of 380 ng of each labelled cDNA sample was fragmented for 30 minutes at 60°C in the [1x] fragmentation buffer (Agilent Technologies, France). Then, the fragmented labelled cDNA samples were diluted into the Agilent hybridization buffer and hybridized onto the probe array for 20h at 60°C in a rotating hybridization oven (Agilent Technologies, France).

An even loop experimental design (Figure II.1) was developed according to variance minimization (Kerr, 2003; Kerr & Churchill, 2001). In this experimental design, every time point was compared to the four others in nine dye-swaps, except for the comparison between the more distant ones: P46 and L40. One dye-swap corresponded to one comparison which meant two slides.



Dye-Swap	Slide	Time point in Cy3	Time point in Cy5
1	S1	P46	P70
	S2	P70	P46
2	S3	P46	P90
	S4	P90	P46
3	S5	P46	P110
	S6	P110	P46
4	S7	P70	P90
	S8	P90	P70
5	S9	P70	P110
	S10	P110	P70
6	S11	P70	L40
	S12	L40	P70
7	S13	P90	P110
	S14	P110	P90
8	S15	P90	L40
	S16	L40	P90
9	S17	P110	L40
	S18	L40	P110

Figure II-1 : Loop experimental design and hybridization table

After the hybridization step, microarrays were washed twice under orbital agitation for 5 minutes at 50°C in a (2X SSC, 0.1% SDS) solution and were successively washed twice for 5 minutes at room temperature in (0.5X SSC, 0.1% SDS) solution, in (0.1X SSC, 0.1% SDS) solution for 5 minutes, in (0.1X SSC) solution 3 times for 5 minutes and in ultra pure water for 1 minute. They were finally dried for 45 seconds in a mini-centrifuge system.

Microarrays were scanned on an Agilent scanner (Agilent Technologies, France) at a 10 µm x 10 µm pixel resolution. Agilent Feature Extraction 9.1.3.1 software (Agilent Technologies, France) was used to quantify mean and median signals for each sample and background estimation following to the GE2_NonAgilent_91 extraction protocol. Extracted data were stored in the BioArray Software Environement (BASE) of SIGENAE (Système d'Information du programme d'Analyse des GENomes des Animaux d'Elevage (SIGENAE) 2008 [http://www.sigenae.org]) and then formatted according to MIAME standards (Brazma *et al.*, 2001) and submitted to the Gene Omnibus (GEO) database as series GSE14008 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=jrqrhioqcywkmtc&acc=GSE14008).

Data were analyzed on R program writer software (http://www.r-project.org/).

Statistical analyses

Three animals were used per developmental stages in order to reduce the impact of individual variability. Data were normalized using the “normalisation” function of the “anapuce”

package of R. It performed an intensity dependant normalization using a global lowess (Yang *et al.*, 2002b) followed by subtraction of the log-ratio median calculated over the values for an entire block from each individual log-ratio in order to correct an eventual print-tip effect.

First, we calculated the average normalized signal for each spot in any time point. The most expressed genes were retained when the corresponding spot signal intensity, in a selective time point, represented more than 10% of the most intense signal observed. Moreover, the genes that were globally highly expressed in every time points were also selected.

In a second data analysis, experimental normalized data extracted from the microarrays of the dye-swaps S1, S4, S7 and S9 (Figure II.1) were statistically analyzed to compare two successive time points, respectively P46 with P70 (P46 / P70), P70 with P90 (P70 / P90), P90 with P110 (P90 / P110) and P110 with L40 (P110/ L40). A differential analysis with a variance mixture modeling (Delmar *et al.*, 2005), based on the "DiffAnalysis" function developed in the "anapuce" package of R, was performed on the normalized data set. This method built groups of genes. Within each group, genes had a homogeneous variance. In order to take into account the multiplicity of tests, *p*-values were adjusted with Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) which controlled the False Discovery Rate (FDR) with a significance threshold fixed at 5%.

In the third analysis, all the data set was analyzed with a linear model including a stage effect. We used the covariance matrix of the parameter estimates to calculate the variance of all contrasts between successive stages $\mu_{i+1} - \mu_i$ and assessed their significance using a *t*-test. *P*-values were corrected using the Benjamini-Hochberg procedure. Genes that vary significantly at least once during pregnancy and lactation were clustered using KMC (K-means clustering) function of the TMeV v4.0 software (Saeed *et al.*, 2003; Soukas *et al.*, 2000). K-means clustering was considered as a good tool since it builds homogeneous clusters, by characterizing each cluster with a central point (its mean), assigning each gene to the closest cluster, and re-calculating the mean of the cluster (Mary-Huard *et al.*, 2006). Parameters used to build the clustering were Euclidean distance and a maximum of 50 iterations.

General slides information and probe annotations

In order to recover GeneBank accession number and human reference for each probe of the slide, we used a web-accessible resource established by SIGENAE (http://public-contigbrowser.sigenae.org:9090/Bos_taurus/index.html) selecting either Human Unigene or SwissProt filters. To annotate the probes in this contig browser (BioMart), Sigenae team first compared the 70mer oligonucleotide sequences spotted on the slide with the latest version of

the bovine genome assembling (Cattle V8 s.bt.8 based on p.bt.5 published in November 2007). If the 70mer oligonucleotide sequence matched perfectly and at a single place with the bovine assembling, the annotation associated with this locus was directly selected to annotate the probe. If the match was imperfect, the 70mer oligonucleotide sequence was compared with all Ensembl transcripts and the retained probe annotation corresponded to the contig number that contained the sequence.

Among the 22,012 oligonucleotide sequences spotted on the glass slides, 416 were control spots and 21,596 corresponded to other sequences. Among these 21,596 sequences, 14,687 (~ 68%) had a human reference accepted by IPA software. About 32% of these 14,687 "IPA genes" were redundant between "Operon" and "Illumina" sets. In total, the oligoarray slides used in this work contained 9,973 different genes analyzed in IPA and about 7,000 probes with Sigenae annotation not accepted by IPA. Taken together, all these nucleotide sequences should represent quite 50% of the whole bovine genome. Only IPA biological functions of which P values were inferior to 5% are presented and discussed in this paper.

Identification of biological functions

For every analysis, significant data were referenced to biological functions with Ingenuity Pathway Analysis software v6.0 (IPA, Ingenuity Systems, Redwood City, CA). The contribution of genes to the identified function was explained with an associated *p*-value, calculated with the right-tailed Fisher's Exact Test. In this test the number of user-specified genes of interest that participated in a given function or pathway were compared to the total number of occurrences of these genes in all functional/pathway annotations stored in the Ingenuity Pathways Knowledge Base. Only significant top IPA molecular and cellular functions and IPA physiological system development and functions are presented and discussed in this article.

Primer design and quantitative real-time PCR

Total RNA (1 µg) was reversed transcribed with 0.5 µg of Oligo(dT)12-18 primer (Invitrogen, France) and SuperScript™ III (Invitrogen, France). The quantitative real time PCR (qPCR) was carried out as described by Bevilacqua et al. (Bevilacqua *et al.*, 2006), using the Sybr green-based fluorescent detection system and the ABI Prism 7900 Sequence Detection System (Applied Biosystems, Foster City, CA) to measure mRNA abundance. A total of 24 genes were validated by qPCR, using three potential reference genes, cyclophilin A, 18S ribosomal RNA and S24 ribosomal protein (*RpS24*), as internal control. As previously

reported (Bevilacqua *et al.*, 2006) *RpS24* messenger has been chosen as internal control. Primer pairs given in Table II.1 were designed using the Primer Express Software v2.0 (Applied Biosystems, France) starting from publicly available caprine and bovine sequences. To avoid genomic DNA amplification, primers were mostly selected on exon-exon junctions from the goat sequences when available or otherwise from the bovine sequences. Primers were purchased from MWG Biotech (France).

All melt curves showed a single amplicon and the specificity of each qPCR assay was confirmed by amplicon size and sequence. Amplification efficiency was measured for each system and differences in gene expression were analyzed using One-way ANOVA and $P<0.01$ was considered statistically significant.

To compare the expression profile of selected genes using qPCR and microarray experiments we first normalized (24S transcript as internal control) individual qPCR data. Relative abundances of target transcripts between the different stages of pregnancy were determined assuming the expression of individual lactation was 1,000. The same procedure was applied to microarray data, except normalization which was performed as described above. Then the mean and standard deviation for each developmental stage were calculated and bar charts were drawn.

Table II-1 : Primers used in this study. Each pair of primers amplifies the target cDNA (amplicon sizes ranging between 54-229 nucleotides). Primers pairs were designed with the Primer Express Software v2.0 (Applied Biosystems) excepts for 24S ribosomal protein primers which were manually designed.

Genes	Primers	Sequence 5'_3'	Amplicon size
αs1-casein	Forward	TCC ACT AGG CAC ACA ATA CAC TGA	61 nt
(*CSN1S1*)	Reverse	GCC AAT GGG ATT AGG GAT GTC	
αs2-casein	Forward	CTG GTT ATG GTT GGA CTG GAAAA	76 nt
(*CSN1S2*)	Reverse	AAC ATG CTG GTT GTA TGA AGT AAA GTG	
β-casein	Forward	GAA AGC CAG AGC CTG ACT CTC A	63 nt
(*CSN2*)	Reverse	CTG GAC CAG AGC CAG AGG AA	
κ-casein	Forward	AGG TGC AAT GAT GAA GAG TTT TTT C	66 nt
(*CSN3*)	Reverse	CCC AAA AAT GGC AGG GTT AA	
α-lactalbumin	Forward	TGG TGC AAA GAC GAC CAG AA	64 nt
(*LALBA*)	Reverse	GGA ACT TGT CAC AGG AGA TGT TAC A	
β-lactoglobulin	Forward	CAT CGT CAC CCA GAC CAT GA	54 nt
(*BLG*)	Reverse	CAA GTC CCC GCC AC	
Lactoferrin	Forward	GAG ACC AAC GGA AGG GTA CCT	59 nt
(*LTF*)	Reverse	TGA GCC CCT CAT TTG CTT TC	
Osteopontin	Forward	CCC AGG AGG AGA GCA AGC ATT	66 nt
(*OPN*)	Reverse	TCT TGG CTG AGT TTG GAA TTT TC	
GlyCAM-1	Forward	CTA CCC TTG GAT CAG AAG AGA CTA CA	69 nt
	Reverse	TCA GTT TTC CTT CTG TGG TGG AT	
JDP-1	Forward	CCC TCC TGT GCT GCC AAC T	91 nt
	Reverse	GAT GGA GGA ATG ATT TGG TAC TCA	
Diacylgylcerol-O-acyltransferase	Forward	GGC GGT CCC CAA CCA	58 nt
(*DGAT1*)	Reverse	GCA GGA GTG GAA GAG CCA GTA	
Stearoyl CoA Desaturase	Forward	TGC TGA CAA CTT ATC TGG ATG C	178 nt
(*SCD*)	Reverse	AAG GAA TC TGC AAA CAG CTA	
Fatty Acid Synthase	Forward	ACA GCC TCT TCC TGT TTG ACG	225 nt
(*FASN*)	Reverse	CTC TGC ACG ATC AGC TCG AC	
Adipophilin	Forward	CCC ACT GTG CTG AGC ACA TT	57 nt
(*ADRP*)	Reverse	GAG TCA GGT TCC GGG CAA T	
FABP-4	Forward	TGG TGC TGG AAT GTG TCA TGA	63 nt
	Reverse	TGG CTT ATG CTC TCT CGT AAA CTC T	
BMP-7	Forward	AAG CAC GAG CTA TAC GTC AGC TT	62 nt
	Reverse	TCG GGT GCG ATG ATC CA	
ID-2	Forward	GAA GGT GAG CAA GAT GGA AAT CC	191 nt
	Reverse	GGA ATT CAG AAG CCT GCA AGG AC	
Elf-5	Forward	TCA AGA CTG TCA CAG TCA TAG TCG AA	65 nt
	Reverse	CTC GCA CAA ATT CCC ATA GAT G	
Butyrophillin	Forward	GAT GGC AGT CTT TCC AAA TTC C	88 nt
(*Btn1A1*)	Reverse	AAA GGG AGC AGA ATC CAG CTT	
Lactadherin	Forward	TGG ATA ATC AGG GCA AGT TCA A	72 nt
(*MFGE-8*)	Reverse	GGT CAA TCT GCA GCC ACT CA	
Lactoperoxidase	Forward	CTC AAC CCT CAC TGG AAT GGA	87 nt
(*LPO*)	Reverse	GTC CCT AAA GGT GAT GAT CTG TAT GA	
Casein kinase 2	Forward	GGA CAT GAC AAT TAT GAT CAG TTG GT	101 nt
(*CNSK-2*)	Reverse	CGT GGA TCT AAT TCA ATG TTG TAT TTG	
Keratin 14	Forward	CCC AGC TCA GCA TGA AAG C	57 nt
(*KRT-14*)	Reverse	AGC GGC CTT TGG TCT CTT C	
Insulin like Growth Factor Binding Protein 5	Forward	CAC AGT AGC CCA CAC GGA TAG A	66 nt
(*IGFBP5*)	Reverse	ATT CCG AGT TGC CTA CCG G	
Lipin 1	Forward	CCC TTC TAT GCT GCT TTT GGA	64 nt
(*LPIN1*)	Reverse	TCG TAC AAG CAA GTG GGA GTG T	
ATP-binding cassette, sub-family G (WHITE), member 2*	Forward	CAT TCC TCG ATA CGG CTA TGC	63 nt
(*ABCG2*)	Reverse	TTT GGG ACA AAA CTT CTG CCC	
Chemokine (C-X-C motif) ligand 9*	Forward	GGG CTT GGA AAC CCT CTT AAA	61 nt
(*CXCL9*)	Reverse	GCA TAA GAG AGG ATG TCA GGC CA	
24S ribosomal protein	Forward	TTT GCC AGC ACC AAC GTT G	66 nt
	Reverse	AAG GAA CGC AAG AAC AGA ATG AA	

* non optimized qPCR system

II.1.3.Results

In order to get a better understanding of the terminal differentiation of the mammary tissues we performed three independent analyses of microarray data. For each we determined the most relevant p-value regarding experiments previously carried out in our laboratory (unpublished results) and statistical methods applied. The three independent analyses were: a) extraction of the most highly expressed genes at each developmental stage, b) identification of genes whose expression varied between two successive time points and c) clustering of genes with same expression profile. In addition, expression profiles of relevant genes identified with microarray analyses were validated by qPCR analysis.

The most highly expressed genes at each developmental stage

To get an overview of genes highly expressed at each developmental stage, we selected the most intense spots whose signal was within 10% of the highest signal. Numbers of probes with high signal intensity were 147, 164, 170, 148, and 130 for P46, P70, P90, P110 and L40, respectively (Table II.2).

Table II-2 : IPA Molecular and Cellular Function of gene sets enriched with most highly expressed genes at each developmental stage

Developmental stage	Number of probes which signal is within 10% of the highest signal	Number of probes with IPA annotation (number of probes without IPA annotation)	Ingenuity Molecular and Cellular function
P46	152	78 (49)	Protein synthesis
			Cell Cycle
			Cell Morphology
			Post-translational Modification
			Cell Death
P70	169	87 (50)	Protein synthesis
			Cell Death
			Cell Cycle
			Cell Morphology
			Post-translational Modification
P90	175	92 (53)	Protein synthesis
			Cell Death
			Cell Cycle
			Cellular Movement
			Cell Morphology
P110	153	82 (45)	Protein synthesis
			Cell Death
			Cell Cycle
			Cell Morphology
			Post-translational Modification
L40	130	79 (40)	Cell Death
			Cell Cycle
			Cell Morphology
			Post-translational Modification
			Cell Signaling

Grey color means functions that are common at every developmental stage

Eighty-two probes were common to all time points (Table II.3). They target 48 different genes with known identifiers that mapped to their corresponding gene object in the Ingenuity Pathways Knowledge Base. These focus genes were overlaid onto a global molecular network built from information contains in the database. They were representative of the following molecular and cellular functions: cell cycle, cell morphology, post-translational modification, cell death and cell signaling. Some of these genes have been shown to be involved in apoptosis such as *Bax* (Bcl-2 associated X-protein), *Bcl-2* (B-cell leukemia / lymphoma 2) and *STAT3* (Signal Transducers and Activators of Transcription 3) that were highly expressed in every developmental stage and known to participate to apoptosis and differentiation of some macrophages. Other genes such as *HMGA1* (High Mobility Group AT-hook 1), *ZAK* (sterile alpha motif and leucine zipper containing kinase AZK) and *EEF1A1* (Eukaryotic translation Elongation Factor 1 α 1) that are regulators of transcription and translation belong also to this dataset. Likewise, *RHOC* (ras homolog gene family, member C) that promotes reorganization of the actin cytoskeleton and regulates cell shape, attachment, and motility give high signal intensities at every developmental stage. The same statement was done for

EGFL7 (EGF-like-domain, multiple 7) which is involved in the formation and maintenance of endothelial integrity (angiogenesis)

This reflects that the goat mammary gland is during pregnancy and lactation a very dynamic organ requiring permanent remodeling and high cell turnover. These biological processes were characterized by high signal intensities in pregnancy and lactation for genes representative of apoptosis, cellular organization and regulators of transcription and translation. In this gene set, the presence of two casein genes (*CSN2* and *CSN3*) was not expected, especially for their high signal intensity at every stage of pregnancy.

In addition, the protein synthesis function here mainly represented by genes coding for ribosomal proteins (RPLs) is systematically associated with every stage of pregnancy. Surprisingly this biological function is not associated with lactation which is however a period of intense milk protein biosynthesis.

Table II-3 : 82 probes with high signal intensity in each developmental stages

Encoded protein	Gene symbol	Probe number	GeneBank accession	P46 signal intensity	P70 signal intensity	P90 signal intensity	P110 signal intensity	L40 signal intensity
Abhydrolase domain containing 6	*ABHD6*	OLIGO_05640	BM363585	70662	64367	55108	63433	106667
Actin, alpha 1, skeletal muscle	*ACTA1*	OLIGO_10461	CR454575	22650	31050	31318	25822	16145
Actin, beta	*ACTB*	OLIGO_10564	AY141970	29377	31095	25831	28848	12223
B-cell CLL/lymphoma 2	*BCL2*	OLIGO_02195	CN436558	69967	63010	54874	63737	100962
BCL2-associated X protein	*BAX*	Bt00002710	CK774062	12806	17433	19299	19890	19124
Beta-2-microglobulin	*B2M*	OLIGO_11407	BF440417	10968	23987	15846	22469	15335
		Bt00007118	CK849269	27367	37569	27802	44716	24612
Casein beta	*CSN2*	OLIGO_12257	S67277	25928	54467	50863	26967	104046
		Bt00007391	CK849815	28103	54410	39221	29941	98833
Casein kappa	*CSN3*	OLIGO_11992	X00565	19374	56202	18506	21580	103328
Cerebellar degeneration-related protein 1, 34kDa	*CDR1*	OLIGO_11327	CR453915	20648	25857	27870	21295	10738
Cytochrome c oxidase subunit 8A (ubiquitous)	*COX8A*	Bt00000316	CK769747	13706	24849	20057	28151	46645
DAZ associated protein 1	*DAZAP1*	OLIGO_11075	EH377586	18023	25615	24865	20848	14649
DEAD (Asp-Glu-Ala-Asp) box polypeptide 46	*DDX46*	OLIGO_10511	CR453981	28377	29293	27632	24707	12204
EGF-like-domain, multiple 7	*EGFL7*	OLIGO_12545	EH169691	14265	13166	13925	14400	13288
eukaryotic translation elongation factor 1 alpha 1	*EEF1A1*	OLIGO_10610	CN433968	34200	41181	47041	39983	29081
		Bt00002530	CK773744	59088	57450	54564	57675	51318
Exosome component 8	*EXOSC8*	Bt00006744	CK848579	18813	17998	20844	19254	18978
Ferritin, heavy polypeptide-like 8	*FTHL8*	OLIGO_07281	CN434013	43078	46543	45420	36365	26042
Glutathione S-transferase pi	*GSTP1*	OLIGO_10473	EH379175	62282	55992	53868	49250	29459
Glycoprotein M6A	*GPM6A*	OLIGO_10501	CR551597	21787	30454	31118	26676	14480
		Bt00006995	CK849066	26402	35516	36524	30949	16814
Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	*GNB2L1*	Bt00007200	CK849439	21161	26347	30153	22825	12102
High mobility group AT-hook 1	*HMGA1*	OLIGO_07163	CR452611	26485	28227	29146	23502	10863
Hypothetical protein LOC339290	*LOC339290*	Bt00000408	CK769920	15820	17547	18194	14132	11625
Integrator complex subunit 3	*INTS3*	Bt00000412	CK769926	32148	28522	32901	28292	88903
Keratin 6A	*KRT6A*	Bt00008172	CK955286	24615	22342	25398	23456	21781
MOB1, Mps One Binder kinase activator-like 2C (yeast)	*MOBKL2C*	Bt00007576	CK943263	13108	15565	16746	17563	14164
Myosin regulatory light chain MRCL3	*MRCL3*	Bt00007144	CK849324	24658	21134	21884	21536	18470
Opsin 1 (cone pigments), medium-wave-sensitive 2	*OPN1MW2*	OLIGO_12108	AF280398	12485	12843	14903	13593	16638
Phospholipase D3	*PLD3*	Bt00003368		36334	36603	33560	34081	50132
Prostaglandin E synthase	*PTGES*	OLIGO_09243		26519	26191	27397	23697	24878
Ran GTPase activating protein 1	*RANGAP1*	Bt00005713	CK846479	24926	24811	28180	30320	25624
Ras homolog gene family, member C	*RHOC*	OLIGO_01134	CR552385	37867	43941	36870	43941	68560
Ribosomal protein L10	*RPL10*	OLIGO_06837	BF045297	21401	25068	25600	22234	13302
		Bt00007021	CK849109	22023	25798	27725	23311	16193

Ribosomal protein L17-like	*HCG 2004593*	Bt00001141	CK771192	29040	33942	34029	28906	19180
		OLIGO_06891	CR454801	34954	40961	39388	35710	20690
Ribosomal protein L23	*RPL23*	OLIGO_10590	CR452075	35538	43368	47155	38707	24053
Ribosomal protein L36a-like	*RPL36AL*	OLIGO_10917	BM363402	20781	24617	22475	20321	13819
Ribosomal protein L4	*RPL4*	OLIGO_07375	BF440439	17824	21460	23852	20112	11577
Ribosomal protein L41	*RPL41*	OLIGO_10576	AW465014	46969	49141	46586	42458	36860
Ribosomal protein L6	*RPL6*	OLIGO_10608	BM363457	18222	25688	26952	20950	11027
		Bt00006929	CK848933	26439	31543	35234	26695	15613
Ribosomal protein S10	*RPS10*	OLIGO_10601	EH165124	11382	18989	19086	14215	12004
Ribosomal protein S11	*RPS11*	OLIGO_10459	DT842423	27674	35839	35115	31161	19139
Ribosomal protein S2	*RPS2*	OLIGO_08626	AW466111	15046	17197	22207	16516	11271
Ribosomal protein S23	*RPS23*	OLIGO_10463	CR454819	27393	34731	34224	27883	16890
		Bt00004836	CK778043	36148	39492	40833	34390	19689
Ribosomal protein S24	*RPS24*	Bt00006435	CK847945	25467	27000	26670	23064	11748
		OLIGO_10620	CR454824	37205	39034	39562	34866	19127
Ribosomal protein S29	*RPS29*	OLIGO_10417	CR452232	21494	29688	26417	21878	12710
Signal transducer and activator of transcription 3 (acute-phase response factor)	*STAT3*	Bt00000335	CK769787	58033	63791	52025	53799	91414
Similar to proline-rich synapse-associated protein 2		OLIGO_09415	CK394152	19122	18117	23367	21015	24599
Solute carrier family 39 (zinc transporter), member 4	*SLC39A4*	OLIGO_12924	BM429756	13997	13915	14830	14033	13852
SRY (sex determining region Y)-box 4	*SOX4*	OLIGO_11229	CN436822	14428	17756	15603	17129	20925
Sterile alpha motif and leucine zipper containing kinase AZK	*ZAK*	OLIGO_10531	CR454290	22383	26670	27027	22517	13504
Survival of motor neuron 1, telomeric	*SMN1*	Bt00002914	CK774453	40931	39579	43211	39350	44344
TAR (HIV-1) RNA binding protein 2	*TARBP2*	Bt00002623	CK773899	12008	15372	16636	16040	16239
Transmembrane protein 28	*TMEM28*	OLIGO_12870	BM253177	11348	13237	14867	14119	17203
Transthyretin (prealbumin, amyloidosis type I)	*TTR*	OLIGO_08138	BF040485	69094	62820	54502	64261	106021
Transthyretin (prealbumin, amyloidosis type I)	*TTR*	OLIGO_10391	BM362443	69794	62999	54424	63265	104161
Tumor protein, translationally-controlled 1	*TPT1*	Bt00002885	CK774391	53279	53546	49114	45089	65618
Ubiquitin A-52 residue ribosomal protein fusion product 1	*UBA52*	OLIGO_06936	CR451887	20428	20667	19977	15421	21285
Ubiquitin fusion degradation 1 like (yeast)	*UFD1L*	OLIGO_00166	CR550896	14174	13814	14244	14338	15029
Ubiquitin-like protein fubi and ribosomal protein S30		OLIGO_10528	CR454463	19068	19766	19561	18193	10836
Upregulated during skeletal muscle growth 5 homolog (mouse)	*USMG5*	OLIGO_10361	CR456038	37038	38378	39564	33452	20194
Chromosome 14 open reading frame 133	*C14ORF133*	OLIGO_10480	CN439899	26068	34910	36416	31874	22422
Chromosome 21 open reading frame 59	*C21ORF59*	OLIGO_01970	CN435801	11960	12170	13508	12567	18473
NA*	NA	OLIGO_09392	AY563835	13143	12643	13387	13077	14382
NA	NA	Bt00005016		13822	16877	15689	16762	25366
NA	NA	Bt00006423		14407	15038	18180	17149	18549
NA	NA	OLIGO_12452		14874	15260	14934	15171	27193
NA	NA	Bt00000685		15241	15509	15505	16056	15919
NA	NA	OLIGO_00634		15613	11445	17330	13317	14186
NA	NA	Bt00007073	CK849184	29534	32697	37322	31176	18124
NA	NA	Bt00006944		32770	30489	36718	24221	18670
NA	NA	OLIGO_11071	DN845867	33908	39402	29851	28752	26125
NA	NA	Bt00006981	DT895654	34707	41433	42679	39567	65584
NA	NA	Bt00002392		43215	38673	41736	37397	57231
NA	NA	Bt00003641		68770	62161	54954	65248	95884
NA	NA	OLIGO_11122		69876	62228	54291	65667	106850

* NA: probes with annotation not accepted by IPA software

Comparison between two successive time points

To evaluate how different were the developmental stages between two successive time points, four differential analyses have been carried out. The first one was between 46 and 70 days of pregnancy (P46/P70), the next one was between 70 and 90 days of pregnancy (P70/P90), then between 90 and 110 days of pregnancy (P90/P110) and the last one between 110 days of pregnancy and 40 days of lactation (P110/L40). The number of genes differentially expressed between two successive time points gave the first quantitative indicator of the differences. They were 2,566 genes differentially expressed between P46 and P70 and 448 genes, between P110 and L40, in comparison to 35 and 30 genes differentially expressed between P70 and P90, and P90 and P110, respectively (Table II.4). This is consistent with the notion that P46, which represents early pregnancy and L40, which corresponds to an established lactation are extreme developmental stages that showed expression profiles signing deep functional recasting and are therefore very different in comparison to the other three stages of pregnancy.

In the comparison P46/P70, the lowest fold change was 1.4 for 210 probes and the highest one was 177 corresponding to the κ-casein (*CSN3*) of which the expression increased between P46 and P70. Conversely, *IGFBP3* gene (insulin-like growth factor binding protein 3) that supervises cell growth by regulating positively apoptosis and myoblast differentiation and by regulating negatively protein phosphorylation and signal transduction, has its expression decreased between P46 and P70. Other genes of which the expression decreased between P46 and P70, were *COL3A1* (collagen, type III, alpha 1), *DCN* (decorin), *OGN* (osteoglycin), *PLXNB3* (plexin B3) that participate in organ morphogenesis. Genes whose expression increased between P46 and P70 were implicated in immune function (*GBP5*: guanylate binding protein 5, *LBP*: lipopolysaccharide binding protein, *CCL4*: chemokine, C-C motif, ligand 4, *IGHA1*: immunoglobulin heavy constant alpha 1 and *LTF*: lactotransferrin), molecular transport (*UBD*: ubiquitin D and *RBP1*: retinol binding protein 1) and cellular movement (*SERPINE2*: serpin peptidase inhibitor, clade E). Thus, as far as the P46/P70 comparison is concerned, genes of organ morphogenesis showed a decreased expression between P46 and P70 whereas the expression of immune response genes is increasing. This statement is the main conclusion arising from the analysis of the 2,566 genes that are differentially expressed from P46 to P70.

Table II-4 : Identification of the IPA biological functions associated to the differential analysis between two successive time points

Comparison	Number of probes differentially expressed (number of different genes)	Stages*	Molecular and cellular function (number of genes)	Physiological system development and function (number of genes)
P46 / P70	3105 (2566)	P46	Gene expression (134)	Organism development (143)
		P=1368	Cellular growth and proliferation (19)	Embryonic development (77)
		G=991	Cell cycle (9)	Organ development (103)
		NA=167	Cellular movement (41)	Nervous system development and function (83)
			Cell morphology (65)	Tissue development (106)
		P70	Cell-to-cell signaling and interaction (54)	Hematological system development and function (56)
		P=1737	Protein trafficking (12)	Immune and lymphatic system development and function (44)
		G=1167	Molecular transport (53)	Immune Response (67)
		NA=241	Cellular movement (37)	Tissue morphology (7)
			Cellular compromise (8)	
P70 / P90	46 (35)	P70 (37)	Molecular transport (4)	Nervous system development and function (1)
		P=37	Cell cycle (4)	Reproductive system development and function (2)
		G=21	Cell death (5)	Tissue development (1)
		NA=6	Drug metabolism (2)	Immune Response (7)
			Lipid metabolism (2)	Hematological system development and function (5)
		P90 (9)	Cell cycle (2)	Reproductive system development and function (2)
		P=9	Cell death (5)	Tumor morphology (2)
		G=8	Cellular compromise (4)	Organ morphology (2)
		NA=0	Small molecule biochemistry (3)	Skeletal and muscular system development and function (2)
			DNA replication, recombination and repair (3)	Connective tissue development and function (3)
P90 / P110	33 (30)	P90 (7)	Cell-to-cell signaling and interaction (2)	Organ morphology (1)
		P=7	Cell morphology (1)	Cardiovascular system development and function (2)
		G=6	Cell death (1)	Hair skin development and function (1)
		NA=1	Cellular assembly and organization (2)	Embryonic development (1)
			Cell signaling (1)	Reproductive system development and function (1)
		P110 (26)	Cell signaling (7)	Hematological system development and function (10)
		P=26	Nucleic acid metabolism (4)	Immune response (9)
		G=17	Small molecule biochemistry (7)	Skeletal and muscular system development and function (4)
		NA=6	Cellular movement (6)	Immune and lymphatic system development and function (10)
			Cell-to-cell signaling and interaction (9)	Cardiovascular system development and function (4)
P110 / L40	541 (448)	P110 (223)	Cellular movement (45)	Immune response (52)
		P=223	Cell-to-cell signaling and interaction (54)	Hematological system development and function (43)
		G=149	Cell death (64)	Tissue morphology (34)
		NA=28	Cell signaling (24)	Immune and lymphatic system development and function (42)
			Molecular transport (25)	Tissue development (29)
		L40 (318)	Lipid metabolism (26)	Digestive system development and function (2)
		P=318	Molecular transport (27)	Immune response (23)
		G=216	Small molecule biochemistry (44)	Connective tissue development and function (6)
		NA=55	Cellular development (6)	Skeletal and muscular system development and function (13)
			Cellular growth and proliferation (67)	Tissue morphology (16)

*P=number of probes, G=number of different genes with IPA annotation, NA=number of probes without IPA annotation

On the other hand, there were few genes differentially expressed between P70 and P90, with only 8 genes of which the expression was increasing. In this comparison, 21 genes showed a decrease in expression with fold changes ranging between 3.5 and 39 for lactotransferrin (*LTF*) and α_{s2}-casein (*CSN1S2*) genes, respectively. The same tendency was observed with genes involved in the immune response including *IFI6* (interferon, alpha-inducible protein 6), *IFI44L* (interferon-induced protein 44-like) and *CXCL10* (chemokine, C-X-C motif, ligand 10) as well as *BoLA-A* which is a component of the bovine MHC. The decreased expression of *CSN3* gene between P70 and P90, revealed in the most highly expressed genes analysis (Table II.3), was confirmed in this comparison.

Again, few genes are differentially expressed between P90 and P110 (n=30). Genes that are involved in the regulation of transcription and protein synthesis (*MLH1*: mutL homolog 1, *SYVN1*: synovial apoptosis inhibitor 1 or synoviolin and *HSPA1B*: heat shock 70 kDa protein 1B) displayed a decreased expression between P90 and P110. Likewise, *FHOD1* (formin homology 2 domain containing 1) and *INHA* (inhibin alpha) that participate to cellular organization showed the same behavior. Genes which expression increased between P90 and P110 were implicated in the immune response: *BoLA-DGA1*, *BoLA-A*, *CXCL9* (chemokine, C-X-C motif, ligand 9), *GBP5*, *MBP* (myelin basic protein) as well as *LTF* and *TPM3* (tropomyosin 3) that participated to cell motility. During this part of pregnancy, fold changes varied between 3.2 and 15 for respectively the guanine nucleotide binding protein (*GNA13*) gene, whose expression decreased, and the major histocompatibility complex (*BoLA-DQA1*) gene, whose expression increased between P90 and P110, reflecting adjustments rather deep changes in gene expression profiling.

Transition from late pregnancy to early established lactation (P110/L40), showed once more that genes involved in immune response (*BoLA-A*, *GBP5*), chemoattractant cytokines promoting T cell adhesion to endothelial cells (*CXCL9* and *CXCL10*) and *VCAM* (vascular cell adhesion molecule 1) which is up-regulated by cytokines in endothelial cells, showed decreasing expression from P110 to L40 (Table II.5). Conversely, genes involved in lipid metabolism and secretion (*LPL*: lipoprotein lipase, *BTN1A1*: butyrophilin, *SCD*: stearoyl-CoA-desaturase and *FABP3*: fatty acid binding protein 3), in molecular transport (*NCALD*: neurocalcin delta, *NDUFB8*: NADH dehydrogenase, ubiquinone, 1 beta subcomplex, 8, 19kDa and *NUP107*: nucleoporin 107kDa), in protein folding (*JDP-1*: DnaJ homolog, subfamily C, member 12) were shown to display an increased expression. The same

expression pattern was observed for milk protein genes: α_{s1}-, α_{s2}- and β-caseins (*CSN1S1, CSN1S2* and *CNS2*) and alpha-lactalbumin (LALBA). The *CSN3* gene follows the same trend which is significant at a slightly higher FDR threshold (6%). Fold changes varied between 1.9 and 359 (fold change for *CNS1S1*). Finally, while biological functions characterizing late pregnancy were immune response and cell adhesion, lactation was described by lipid metabolism, molecular transport and milk protein synthesis and protein folding.

Table II-5 : Top 25 of up-regulated and down-regulated transcripts in the goat mammary tissue at 40 days of lactation compared to 110 days of pregnancy (adjusted P-value<0.05)

Probe number*	GeneBank accession	Gene symbol	Encoded protein	Log Ratio	Fold change
Up-regulated					
Bt00006833	CK848760	CSN1S1	casein alpha s1	8,5	359,8
OLIGO_12085	EE974964	BTN1A1	butyrophilin, subfamily 1, member A1	6,4	84,8
OLIGO_02149	CN436503	NCALD	neurocalcin delta	6,0	61,9
OLIGO_12034	J05147	LALBA	lactalbumin, alpha	5,8	55,9
Bt00007030	CK849125	LALBA	lactalbumin, alpha	5,7	52,2
OLIGO_12366	EH125902	CSN1S2A	casein alpha s2-like A	5,5	46,8
Bt00003265	CK775070	DNAJB11	DnaJ (Hsp40) homolog, subfamily B, member 11	5,3	39,3
OLIGO_10660	BM363051	FABP3	fatty acid binding protein 3	5,3	38,4
Bt00006728	CK848552	FABP3	fatty acid binding protein 3	5,2	36,4
Bt00007007	CK849087	BTN1A1	butyrophilin, subfamily 1, member A1	5,0	33,0
Bt00007059	CK849167	DNAJC12	DnaJ (Hsp40) homolog, subfamily C, member 12	4,9	30,4
Bt00006984	CK849042	NUP107	nucleoporin 107kDa	4,9	29,6
Bt00002602	CK773866	NDUFB8	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa	4,8	28,8
Bt00007314	CK849661	LPL	lipoprotein lipase	4,8	27,4
Bt00007029	NM_173959	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	4,8	26,9
OLIGO_12092	NM_174528	CSN1S2	casein alpha-S2	4,6	24,4
Bt00008085	CK953877	CCR3	chemokine (C-C motif) receptor 3	4,5	23,3
Bt00002657	CK773964	C9	complement component 9	4,5	22,2
OLIGO_11838	NM_173933	LPO	lactoperoxidase	4,4	20,7
OLIGO_11285	NM_173959	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	4,4	20,5
OLIGO_09278	BF041126	DNAJC12	DnaJ (Hsp40) homolog, subfamily C, member 12	4,3	19,7
Bt00006966	CK848999	MED10	mediator complex subunit 10	4,3	19,3
OLIGO_12148	X81699	SLC34A2	solute carrier family 34 (sodium phosphate), member 2	4,2	19,0
Bt00004167	CK776646	PTPN11	protein tyrosine phosphatase, non-receptor type 11	4,2	18,6
Bt00007394	CK849822	NOP5/NOP58	nucleolar protein NOP5/NOP58	4,2	18,5
Down-regulated					
OLIGO_11084	CR452945	COL1A1	collagen, type I, alpha 1	-2,3	-5,0
OLIGO_06802	CN440096	STMN1	stathmin 1/oncoprotein 18	-2,4	-5,1
Bt00006868	CK848824	CLU	clusterin	-2,4	-5,3
Bt00007372	CK849770	STAT1	signal transducer and activator of transcription 1, 91kDa	-2,4	-5,3
Bt00007230	CK849494	FCGR2A	Fc fragment of IgG, low affinity IIa, receptor (CD32)	-2,4	-5,5
Bt00007018	CK849105	BoLA-A	major histocompatibility complex, class I, A	-2,5	-5,5
Bt00007214	CK849467	BoLA-A	major histocompatibility complex, class I, A	-2,5	-5,5
Bt00003324	CK775170	PSMB10	proteasome (prosome, macropain) subunit, beta type, 10	-2,5	-5,5
Bt00007310	CK849654	CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	-2,5	-5,7
OLIGO_07038	CN435237	OLFML3	olfactomedin-like 3	-2,5	-5,7
OLIGO_08153	CN441870	VCAM1	vascular cell adhesion molecule 1	-2,6	-6,2
OLIGO_10553	CR552676	COL3A1	collagen, type III, alpha 1	-2,7	-6,3
Bt00006666	CK848411	COL3A1	collagen, type III, alpha 1	-2,7	-6,5
OLIGO_12360	AV614702	BoLA-A	major histocompatibility complex, class I, A	-2,8	-6,8
OLIGO_11680	BM362485	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9	-2,9	-7,4
OLIGO_10329	EH374936	BoLA-A	major histocompatibility complex, class I, A	-3,0	-7,8
OLIGO_05858	CV798732	GVIN1	GTPase, very large interferon inducible 1	-3,0	-7,9
OLIGO_09615	BM363744	BoLA-A	major histocompatibility complex, class I, A	-3,0	-7,9
OLIGO_05563	BM362452	CXCL9	chemokine (C-X-C motif) ligand 9	-3,0	-8,1
OLIGO_12484	CK955893	CXCL10	chemokine (C-X-C motif) ligand 10	-3,1	-8,8
Bt00007863	CK949281	GBP1	guanylate binding protein 1	-3,2	-9,4
OLIGO_10483	CR452494	COL1A2	collagen, type I, alpha 2	-3,4	-10,7
Bt00007938	CK950779	UBD	ubiquitin D	-3,7	-12,9
Bt00008202	CK955893	CXCL10	chemokine (C-X-C motif) ligand 10	-3,7	-13,2
OLIGO_05270	BF440277	UBD	ubiquitin D	-4,1	-17,7

* number BtX corresponds to Bovine Genome Oligoset V1.1 (Operon Biotechnologies) ; number OLIGO_X corresponds to NCBI GEO reference GPL2853

Clustering

In order to have the best biological characterization of each gene expression profile, we performed a clustering on the genes that varied significantly at least once in the linear model analysis. In this analysis, the t-test applied to the contrasts between the estimated parameters

allowed to conclude that 2,330 probes corresponding to 1,696 different genes with IPA annotation varied significantly (threshold set to 0.15) at least once in our study. We voluntary fixed the significance threshold to 0.15 to recover the maximum of genes of interest such as genes coding for milk proteins for which qPCR quantification revealed significant variations. These variations were not significant under 0.15 threshold as the fluorescence signal intensity on microarrays saturated for the lactating stage, thus minimizing the differences between stages. The significant threshold chosen here was the best compromise between the number of genes for the biological interpretation and the estimated error.

About 73% of these 2,330 probes showed a fluorescence signal of which the intensity was inferior to 10 times the background signal, 2% were within 10% of the highest signal of the dataset and 25% had a middle signal. All of these genes were divided into 19 clusters according to their expression profile (Figure II.2). We remind here that the clustering was done on the contrasts of the estimated parameters which meant the difference between one stage and the previous one ($\theta_i = \mu_{i+1} - \mu_i$). As the repartition was not completely segregated, which can happen with this clustering method, the lists of genes belonging to clusters with a same profile were merged for IPA functional analysis into metaclusters: merging of cluster 1 with clusters 9 and 12 (metacluster A); merging of clusters 5 and 7 (metacluster B); merging of clusters 6 and 8 (metacluster C); merging of cluster 11 with clusters 14 and 16 (metacluster D); merging of clusters 13 and 15 (metacluster E). Biological functions determined by IPA software associated with each metacluster, or cluster, are given in Table II.6.

In the following text, no precise description was given for clusters 3 and 10 and metacluster C. Indeed, for these groups biological interpretation was not evident as biological function given by IPA were in the limit of significance and genes belonging to these groups were not exclusive in one biological function.

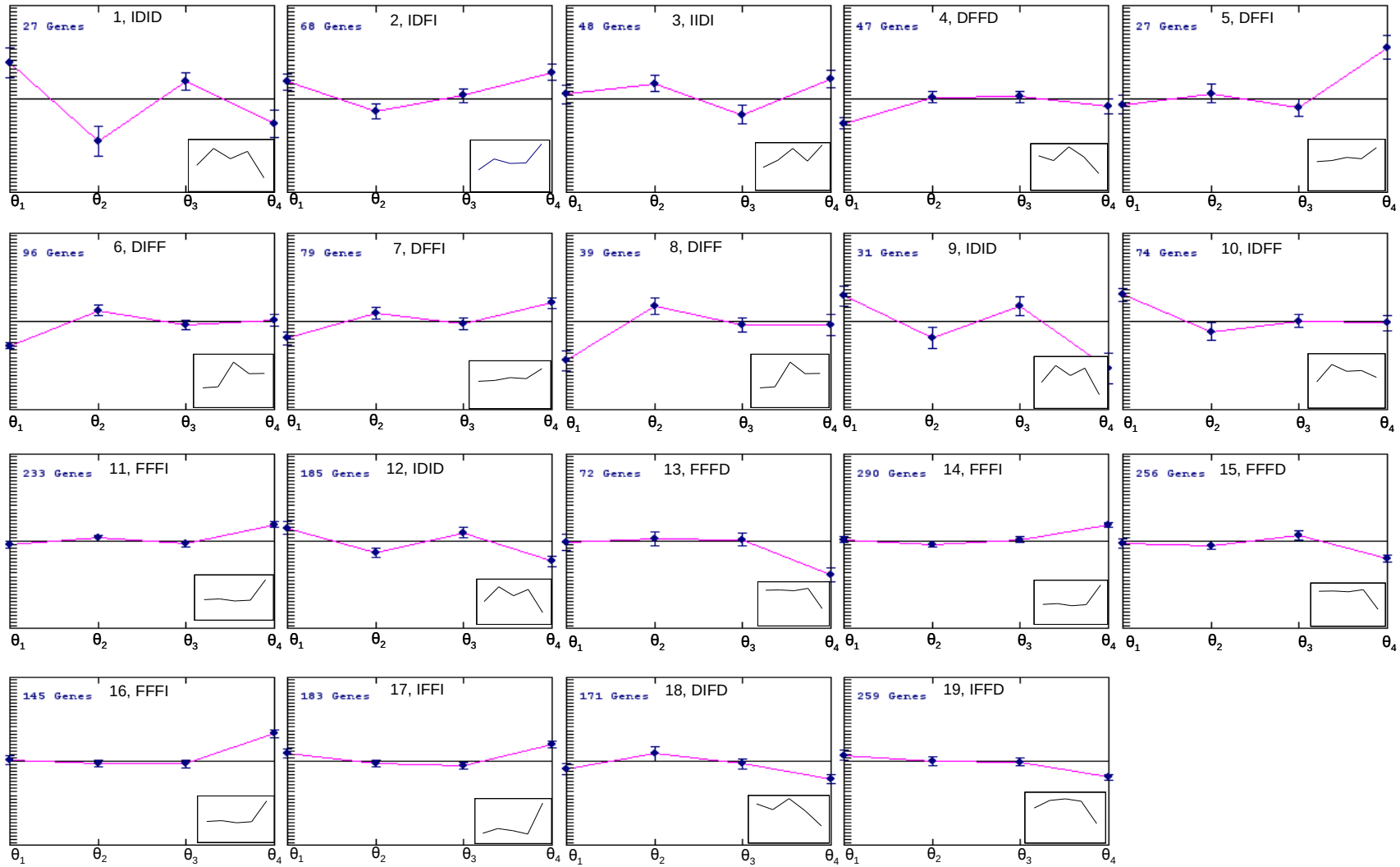


Figure II-2 : K-means clustering allocating into 19 clusters the 2,330 probes equivalent to 1,696 genes that varied significantly at least once
I: Increase, D: Decrease, F: Flat as proposed by Rudolph *et al.*, (2007). θ is the difference between one developmental stage and the previous one. If θ is positive the gene expression goes up, if θ is negative the gene expression goes down and if θ is null the gene expression doesn't vary.In the corner box the gene expression profile drawn from the normalized signal intensity.

Table II-6 : IPA biological functions associated to each cluster

Cluster	Number of probes	Number of annotation recognized by ingenuity	IPA Molecular and Cellular function (number of genes)	IPA Physiological System Development and function (number of genes)	IPA Canonical Pathway (ratio)
1, 9, 12	243	218	Cell death (49) Cell-to-cell signaling and interaction (41) Gene expression (13) Cellular development (22) DNA replication, recombination, and repair (12)	Immune response (50) Hematological system development and function (40) Immune and lymphatic system development and function (40) Tissue morphology (26) Embryonic development (11)	Interferon signaling (9/29) Protein ubiquitination pathway (14/203) Antigen presentation pathway (6/39) Hepatic fibrosis/ hepatic stellate cell activation (6/131) D-glutamine and D-glutamate metabolism (2/27)
2	68	62	Molecular transport (8) Protein trafficking (5) Amino acid metabolism (3) Cell death (1) Cell morphology (2)	Embryonic development (1) Tissue development (1) Organ morphology (1) Organismal development (1) Connective tissue development and function (1)	Cell cycle: G1/S checkpoint regulation (2/60) Glutathione metabolism (2/104) Taurine and hypotaurine metabolism (1/47) Mitochondrial dysfunction (2/165) Arachidonic acid metabolism (2/211)
3	48	40	Cell morphology (6) Amino acid metabolism (2) Carbohydrate metabolism (3) Cell death (5) Cellular function and maintenance (6)	Tissue morphology (4) Cardiovascular system development and function (3) Digestive system development and function (1) Endocrine system development and function (3) Organ morphology (6)	Fatty acid biosynthesis (1/49) G-protein coupled receptor signaling (2/199) Keratan sulfate biosynthesis (1/49) Urea cycle and metabolism of amino groups (1/80) N-glycan biosynthesis (1/87)
4	47	43	Gene expression (3) Carbohydrate metabolism (2) Cell cycle (1) Cell death (9) Cell morphology (3)	Cardiovascular system development and function (2) Endocrine system development and function (2) Hair and skin development and function (1) Hematological system development and function (2) Immune and lymphatic system development and function (2)	Glycosphingolipid biosynthesis - Lactoseries (1/28) O-glycan biosynthesis (1/43) Glycosphingolipid biosynthesis - Globoseries (1/41) Glycosphingolipid biosynthesis - Ganglioseries (1/58) Keratan sulfate biosynthesis (1/49)
5, 7	106	94	Lipid metabolism (8) Nucleic acid metabolism (4) Small molecule biochemistry (13) Post-translational modification (3) Cell death (16)	Hematological system development and function (3) Cardiovascular system development and function (4) Digestive system development and function (4) Embryonic development (3) Endocrine system development and function (3)	Fatty acid biosynthesis (2/49) LXR/RXR activation (3/81) Aryl hydocarbon receptor signaling (3/152) Pyruvate metabolism (2/145) TGF-beta signaling (2/83)
6, 8	135	114	Cell signaling (14) DNA replication, recombinatio and repair (2) Nucleic acid metabolism (4) Small molecule biochemistry (13) Gene expression (20)	Organismal development (15) Behavior (8) Cardiovascular system development and function (9) Nervous system development and function (14) Tissue morphology (7)	VDR/RXR activation (3/80) ERK/MAPK signaling (4/226) Inositol phosphate metabolism (3/173) Wnt/beta-catenin signaling (3/165) cAMP-mediated signaling (3/159)
10	74	63	Carbohydrate metabolism (5) Small molecule biochemistry (8) Cellular development (3)	Embryonic development (5) Tissue morphology (6) Tumor morphology (3)	Oxidative phosphorylation (7/158) Ubiquinone biosynthesis (4/105) Mitochondrial dysfunction (5/165)

			Cellular growth and proliferation (3) Molecular transport (7)	Cardiovascular system development and function (3) Organ development (4)	Valine, leucine and isoleucine degradation (3/107) Citrate cycle (2/59)
11, 14, 16	668	551	Cell death (96) Cellular response to therapeutics (4) Cellular function and maintenance (22) Cellular movement (57) Lipid metabolism (47)	Renal and urological system development and function (9) Cardiovascular system development and function (25) Tissue development (30) Connective tissue development and function (33) Hair and skin development and function (14)	LXR/RXR activation (9/81) Acute phase response signaling (12/172) Hepatic fibrosis / Hepatic Stellate Cell Activation (10/131) LPS / IL-1 meditaed inhibition of RXR function (13/195) EGF Signaling (6/47)
13, 15	328	287	Cellular development (44) Cell cycle (34) Cell morphology (43) Gene expression (38) Cellular growth and proliferation (70)	Immune response (34) Hematological system development and function (34) Immune and lymphatic system development and function (33) Tissue morphology (30) Connective tissue development and function (6)	Antigen presentation pathway (4/39) IL-4 Signaling (5/68) T cell receptor signaling (6/102) B cell receptor signaling (7/148) Glucocorticoid receptor signaling (9/265)
17	183	153	Carbohydrate metabolism (10) Nucleic acid metabolism (7) Small molecule biochemistry (21) Cellular assembly and organization (9) Amino acid metabolism (8)	Skeletal and muscular system development and function (6) Hematological system development and function (11) Immune response (10) Cardiovascular system development and function (7) Connective tissue development and function (1)	N-glycan biosynthesis (3/87) Integrin signalling (5/192) Calcium signalling (4/204) Arginine and proline metabolism (3/178) Urea cycle and metabolism of amino groups (2/)
18	171	158	RNA Post-transcriptional modification (8) Gene expression (13) Cellular development (12) Cell morphology (14) Cellular growth and proliferation (36)	Connective tissue development and function (13) Skeletal and muscular system development and function (7) Hematological system development and function (10) Immune response (6) Immune and lymphatic system development and function (10)	p53 signaling (5/87) VDR/RXR activation (3/80) Wnt/ beta-catenin signaling (4/165) Hepatic fibrosis / Hepatic stellate cell activation (3/131) Huntington's disease signaling (4/232)
19	259	229	Cellular growth and proliferation (36) Protein synthesis (15) Cell-to-cell signaling and interaction (8) Carbohydrate metabolism (5) Cellular development (13)	Cardiovascular system development and function (3) Immune and lymphatic system development and function (13) Tissue development (6) Digestive system development and function (3) Skeletal and muscular system development and function (9)	Protein ubiquitination pathway (9/203) Estrogen receptot signaling (5/118) Selenoamino acid metabolism (3/69) Methionine metabolism (3/76) Synthesis and degradation of ketone bodies (2/19)

Metacluster A (clusters 1, 9 and 12) gather 218 genes showing an expression profile of IDID. The most relevant genes of these clusters were *CXCL10, CXCR4* (chemokine, C-X-C motif, receptor 4), *IL18* (interleukin 18), *IL2RG* (interleukin 2 receptor, gamma), *IRF8* (interferon regulatory factor 8), *IFITM3* (interferon induced transmembrane protein 3), a class of proteins mediating several cellular processes such as homotypic cell adhesion functions of interferons and cellular anti-proliferative activities, *IFNAR1* (interferon, alpha, beta and omega receptor 1) that are coding for cytokines or cytokine receptors and *BoLA-A, BoLA-DRB1* that encode bovine MHC (class I and II, respectively) proteins. These genes were representative of immune response and immune and lymphatic system development and function. The IPA canonical pathways associated to the genes occurring in these clusters were interferon signaling and antigen presentation pathways.

In the cluster 2 there were 62 genes that showed an expression profiles of IFFI (cluster 2, Figure II.2). Some of these genes were *AP3S1* (adaptor-related protein complex 3, sigma 1 subunit), *ERP29* (endoplasmic reticulum protein 29), *RAB13* (member RAS oncogene family), *SEC22B* (vesicle trafficking protein homolog B), *SLC36A1* (solute carrier family 36) which encode proteins that participate to molecular transport and protein trafficking.

Cluster 4 contains 43 genes displaying expression profiles of DFFD. One representative gene of this cluster is adiponectin (*ADIPOQ*) which is mainly expressed in and secreted by the adipose tissue.

94 genes belong to the metacluster B (clusters 5 and 7) and show expression profiles of DFFI. A significant proportion of the genes in this metacluster (n=44) are involved in lipid and nucleic acid metabolism, small molecule biochemistry, post-translational modification, and cell death. One of the canonical pathways associated with these clusters is fatty acid biosynthesis represented by Fatty Acid Synthase (*FAS*) and Acyl-CoA Carboxylase (*ACACA*) genes.

Clusters 11, 14 and 16 (metacluster D), include genes (n=551) whose expression was of FFFI profile. A significant number of these genes are involved in cellular death (n=96) and lipid metabolism and secretion (n=47). Genes participating to lipid metabolism and secretion are exemplified by *ABCG2* (ATP-binding cassette, sub-family G, member 2), *ACACA*, *LPL*, *SDC2* (syndecan 2), *FABP2* (fatty acid binding protein 2), *LPIN1* (lipin 1), *AGPAT1* (1-acylglycerol-3-phosphate O-acyltransferase 1), *ACADSB* (acyl-Coenzyme A dehydrogenase, short/branched chain), *ACSL1* (acyl-CoA synthetase long-chain family member 1), *ALDH3B2* (aldehyde dehydrogenase 3 family, member B2), *DPAGT1* (dolichyl-phosphate N-acetylglucosaminephosphotransferase 1), *HDLBP* (high density lipoprotein binding protein).

Genes representative of cellular death, *Bcl-2*, of which the anti-apoptotic function is well described, and *STAT3*, share the same expression pattern.

On the microarray there are 3 probes corresponding to *ACACA* gene. Two of them where directed in two metaclusters (B and D) underlying the proximity between these two metaclusters. The last probe does not appear in the 19 clusters since its variability was very close to the threshold of significance.

Metacluster E (clusters 13 and 15) brings together 287 genes down regulated in lactation (FFFD). This metacluster is associated with the following biological functions: immune response, immune and lymphatic system development, which are allocated at the molecular and cellular level to cellular development, cell cycle, cellular growth and proliferation. Canonical pathways associated with this profile were antigen presentation pathway, IL-4 signaling, T-cell and B-cell receptor signaling. Some genes representative were *CD74*, *BoLA-A*, *BoLA-DMB* and *BoLA-DRB1* that belong to the bovine MHC, *IL2RG*, *IL4R* (interleukin 4 receptor) that are interleukin receptors, and *NFATC2* (nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2) which is a positive regulator of transcription.

Cluster 17 contained 153 genes which expression profile was IFFI. Some of these genes were *UGP2* (UDP-glucose pyrophosphorylase) and *B4GALT1* (UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 1) that contribute to carbohydrate metabolism, including lactose synthesis. In this cluster there were also genes encoding milk proteins such as α_{s1}-casein (*CSN1S1*) and β-casein (*CSN2*).

Cluster 18 contained 158 genes showing expression profiles of DIFD and was characterized by RNA-post transcriptional modification and gene expression in terms of biological functions. In this cluster there were genes coding for splicing proteins as *SF3B3* (splicing factor 3b, subunit 3), *SFRS4* (splicing factor, arginine/serine-rich 4), *SFRS7* (splicing factor, arginine/serine-rich 7), *SFRS12* (splicing factor, arginine/serine-rich 12), *SFRS* (splicing factor, arginine/serine-rich) or DNA replication or transcription factors such as *HMGB1*, *MED1* (mediator complex subunit 1), *SSRP1* (structure specific recognition protein 1), *WT1* (Wilms tumor 1).

Finally, cluster 19 showed 229 genes which expression profile was IFFD. These genes were representative of cellular growth and proliferation and protein synthesis. Genes of this cluster belong to several biological functions. One gene of this cluster encodes the local factor C/EBPβ which is involved in the cell fate decision leading to progesterone receptor expression and then affects cell proliferation through a paracrine mechanism (Hennighausen & Robinson, 2001).

qPCR results

Some results provided by the microarray analysis were validated by comparison with qPCR performed on the same mammary tissue samples. In addition, since specific genes of milk protein biosynthesis, milk fat globule biogenesis or mammary tissue development were not revealed by this microarray global analysis due to experimental issues (signal saturation for several probes) their expression profile were specifically analyzed by qPCR and then compared to their pattern arising from the microarray analysis. A set of 19 PCR systems was designed to analyze the expression of genes involved in the biosynthesis and secretion of lipids (*DGAT1*, *SCD*, *FAS*, *ADRP*, *FABP4*, *MFGE8*, *BTN1A1*), of milk proteins (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *LALBA*, *LTF*, *LPO*), of mammary tissue development (*BMP-7*, *ID-2*, *OPN*) as well as ubiquitous genes (*JDP-1*, *CSNK2A2*). In addition, 4 genes relevant for the mammary function but not represented on microarrays (*KRT14*, *GLYCAM1*, *BLG*, *Elf5*), and 4 genes representative of main clustering expression profiles (*CXCL9*, *ABCG2*, *IGFBP5*, *LPIN1*) were also analyzed.

The expression profiles from microarrays show the same trends as qPCR for most of the genes investigated with a global correlation coefficient close to 0.82 (Figures II.3 and II.4). Nevertheless, some genes do not show acceptable correlation between qPCR and microarray expression profiles as for example *DGAT1*, *CSN1S2*, *CSN2* and *CSN3* (Figure II.3). For the *DGAT1* gene (Figure II.3A), the signal intensity was weak in every stage (under 10 times the smallest signal intensity of the data set). As the complete DGAT1 mRNA sequence is incomplete for *Capra hircus* and *Ovis aries* we compared the nucleotide sequence of the DGAT1 microarray oligonucleotide probe with the DGAT1 mRNA sequence of *Bos taurus* (NCBI: NM_174693), *Homo sapiens* (NCBI: NM_012079), *Mus musculus* (NCBI: NM_010046), *Sus scrofa* (NCBI: AY093657) and *Bubalus bubalis* (NCBI: DQ886485). The number of mismatches was respectively 0, 30, 30, 19 and 4. This region seems highly variable between species, thus suggesting the occurrence of mismatches between the microarray probe and the *Capra hircus* mRNA sequence, which can explain the weak signal intensity and the absence of difference between developmental stages with the microarrays. For the casein genes (Figure II.3B), relative abundances between L40 and P70 were higher with qPCR analysis than with microarray analysis. This can be due to saturation in the fluorescence signal on microarray leading to an underestimation of differences.

There are two gene expression profiles for the lipid biosynthesis genes (Figure II.3A) with genes highly expressed only in lactation (*BTN1A1*, *SCD*, *MFGE-8*) and genes that are expressed at every stage of pregnancy (*FASN*, *ADRP*, *FABP4*). The milk protein genes can be

divided into 3 expression profiles (Figure II.3B) with genes that are expressed at mid-pregnancy and highly expressed in lactation (*CSN2*, *CSN1S2*, *CSN3* and *LALBA*), genes that are highly expressed only in lactation (*CSN1S1* and *LPO*) and genes that are highly expressed at mid-pregnancy and weakly expressed in lactation (*OPN* and *LTF*). Genes known for their implication in the mammary tissue development (*BMP-7* and *ID-2*) were expressed at every developmental stage (Figure II.3C) and *CSNK2A2* displays the same expression profile.
To validate expression profile obtained by clustering, we have selected some key genes in the main expression profiles. Indeed, *LPIN1*, *ABCG2* and *IGFBP5* showed up-regulation in lactation and *CXCL9* was up-regulated at P70 and P110 (Figure II.4).

Other expression profile of genes with key interest in the mammary tissues functionality but not present on the microarray were determined by qPCR (Figure II.5). *JDP-1*, *BLG*, *GLYCAM1* and *Elf5* showed high expression in lactation with a noteworthy expression of *Elf5* at P70. *KRT14* was equally expressed at every developmental stage.

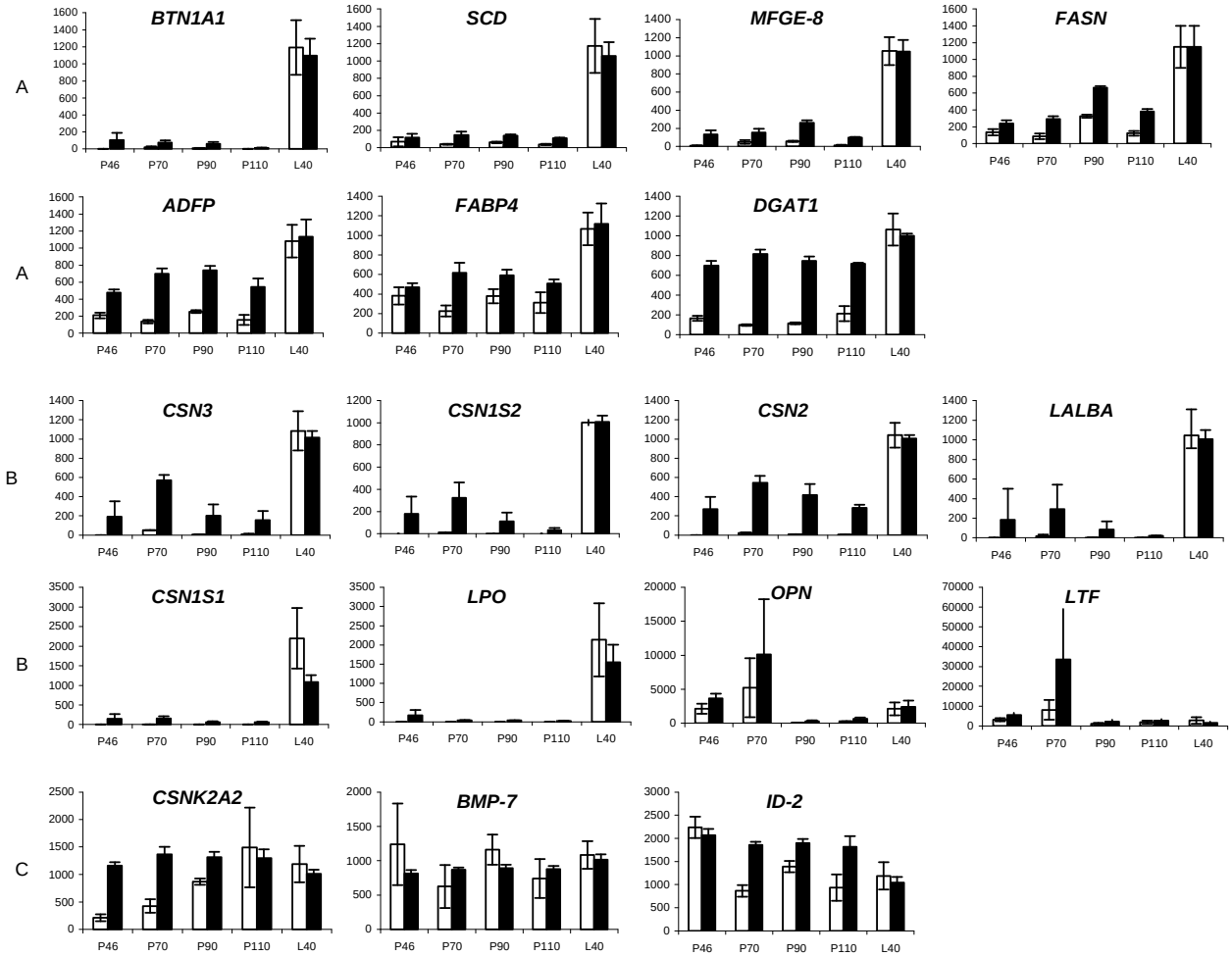


Figure II-3 : Gene expression profiles obtained by qPCR and microarrays.

A: lipid biosynthesis genes, B: milk protein genes, C: mammary setting and ubiquitary genes. *BTN1A1*: Butyrophillin, *SCD*: Stearoyl CoA Desaturase, *MFGE-8*: Milk Fat Globule-EGF factor 8, *FASN*: Fatty Acid Synthase, *ADFP*: Adipophillin, *FABP4*: Fatty Acid Binding Protein 4, *DGAT1*: DiacylGlycerol AcyTransferase 1, *CSN3*: Casein κ, *CSN1S2*: casein α_{S2}, *CSN2*: Casein β, *LALBA*: α-lactalbumin, *CSN1S1*: Casein α_{S1}, *LPO*: Lactoperoxydase, *OPN*: Oesteopontin, *LTF*: Lactoferrin, *CSNK2A2*: Casein Kinase 2, *BMP-7*: Bone Morphogenetic Protein 7, *ID-2*: Inhibitor of DNA binding 2. White (qPCR) and black (microarray) bar represent the relative abundancies of target transcripts between the different stages of pregnancy, determined assuming the expression of individual lactation was 1,000.

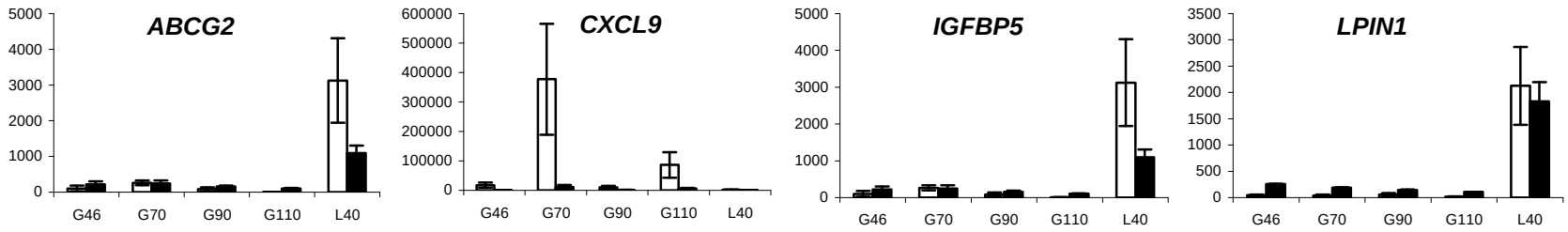


Figure II-4 : Gene expression profiles obtained by qPCR and microarrays of key genes in clustering.

***ABCG2*: ATP-binding cassette sub-family G (White), member 2, *CXCL9*: Chemokine (C-X-C motif) ligand 9, *IGFBP5*: Insulin-like Growth Factor Binding Protein 5, *LPIN1*: Lipin 1. White (qPCR) and black (microarray) bar represent the relative abundancies of target transcripts between the different stages of pregnancy, determined assuming the expression of individual lactation was 1,000.**

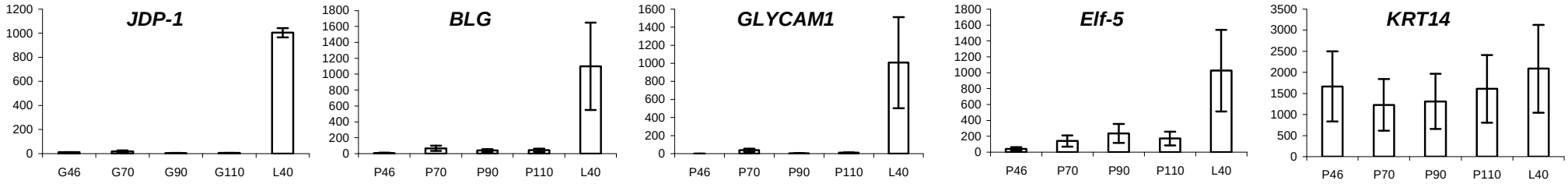


Figure II-5 : Gene expression profiles obtained by qPCR.

***BLG*: β-lactoglobulin, *GLYCAM1*: glycosylation dependant cell adhesion molecule 1, *Elf-5*: E74-like factor 5, *KRT14*: Keratin 14. White bar represent the relative abundances of target transcripts between the different stages of pregnancy, determined assuming the expression of individual lactation was 1,000.**

II.1.4.Discussion

This study is the first global gene expression profiling of differentiation, designed in ruminants, to follow changes occurring in the ontogeny of the mammary gland during pregnancy. We have chosen the goat species as it was a smaller and cheaper ruminant model than bovine. As pregnancy is about 150 days long for goats, P46 represents the end of the first third of the goat pregnancy time and P70 represents mid pregnancy. Three remarkable results describing the goat mammary gland terminal differentiation have to be mentioned. The first one, refers to the first half of pregnancy, which was characterized by a decrease in the expression genes usually expressed in adipocytes whereas genes representative of a differentiated mammary epithelium such as genes coding for immune function and milk proteins had their expression increased. The second main result is the switch between late pregnancy (P110) and lactation showing the decreased expression of immune function encoding genes and the increased expression of genes coding for milk components (especially lipids and lactose) biosynthesis and secretion. The third outstanding result is the pointed out increase expression of genes encoding both proteins of the innate and specific immunity at mid-pregnancy (P70) and at the last third of pregnancy (P110).

First half of pregnancy, change in cell-types

It seems that many changes occurred in the goat mammary gland during the first half of pregnancy. Indeed, our data suggest that the goat mammary gland move a decrease in the ratio of stroma (mainly composed of adipocytes and fibroblasts) to secretory tissue (composed of epithelial and myoepithelial cells) as witnessed by the reduction in the expression of adipocyte markers, *e.g.* adiponectin (Hu *et al.*, 1996), belonging to cluster 4 or fibroblasts markers, *e.g.* FGF-10 (Pedchenko & Imagawa, 2000). A number of adipocyte markers (Master *et al.*, 2002), including PPARγ, also showed a decrease in their expression, between P46 and P70.

Moreover, the expression profiles of *VEGF* and *SREBF1*, whose expression decreased, *Elf5* and cyclin D1 (*CCND1*) both whose expression increased between P46 and P70, which are known to be involved in mice lobulo-alveolar development and differentiation (Anderson *et al.*, 2007), suggest that the goat mammary tissue follows the same pattern of differentiation process. In addition, *ID2* which has been shown to be required for lobulo-alvelolar development (Mori *et al.*, 2000), is significantly expressed during pregnancy in the goat mammary tissue. In mice, it has been well described that the switch between the first third and

mid-pregnancy is characterized by a lobulo-alveolar development (Anderson *et al.*, 2007; Hennighausen & Robinson, 1998; Pitelka *et al.*, 1973; Robinson *et al.*, 1995). Moreover, research works of Wang and coll. (2006) shows that the variation in gene expression is more dependant of the change of cell composition in the sample than a variation in expression of this gene in one type of cell (Wang *et al.*, 2006). Hence, we can hypothesize from our result that between P46 and P70, the goat mammary gland is changing from a stromal tissue made of adipocytes and fibroblasts to an organized lobulo-alveolar tissue with an epithelium internally coating the alveoli in which milk will be synthesized and secreted after parturition. This hypothesis is validated by the only existing morphological study done in 1979 by Swanson and Poffenbarger. They observed that the bovine mammary tissue has already alveolar structures at mid-pregnancy (5 months in their study).

Epithelium differentiation and activation of secretion

Data produced from the different analyses in our study revealed that genes encoding κ-casein and to a lesser extent α_{s2}- and β-caseins, α-lactalbumin and β-lactoglobulin are expressed at mid-pregnancy in the goat mammary tissue. Moreover, we show that Elf5, a member of the Ets transcritpion factor family (epithelium-specific subclass) which regulates a number of epithelium-specific genes found in tissues containing glandular epithelium, is also expressed at mid-pregnancy. In the mouse mammary tissue, milk protein synthesis is regulated by the presence of two transcription factors: Stat5 and Elf5 (Anderson *et al.*, 2007). The presence of Elf5 and milk protein gene transcripts in the goat mammary tissue at mid-pregnancy meets the theory developed for mouse mammary tissue according to which at mid-pregnancy there are sufficiently mammary epithelial cells differentiated to express markers of milk synthesis (Robinson *et al.*, 1995). Actually, milk protein genes are slightly expressed in the pregnant goat mammary tissue as previously shown, using an immuno-histochemistry approach in the pregnant mouse mammary tissue (Kanazawa & Kohmoto, 2002). Semiquantitative Western blotting analysis of α_{s1}- and β-caseins in epithelial cell fractions from mouse mammary glands showed that intra-epithelial caseins are present at a trace level during initial stages of pregnancy (between days 0 and 4), declined to lower levels during mid-pregnancy (days 6 to 8), and then rose to high levels during the last third of pregnancy (day 14). Conversely, at the messenger level (our qPCR data) we have recorded in the goat a trend to a slight increase in expression at mid pregnancy for casein encoding genes, especially *CSN3* (κ-casein). However, it is worth mentioning that casein synthesis can substantially be regulated at the

level of translation (Choi *et al.*, 2004). Nevertheless, for both goat and mice, it has to be mentioned that the expression of milk protein genes remains very limited during pregnancy as compared with the outburst occurring at the onset of lactation (more than 1,000-fold from qPCR results), when the mammary gland is supposed to be totally differentiated and produce huge amounts of milk. The fold change obtained with the microarray technology is more limited (359-fold) though still consequent with α_{s1}-casein transcripts between P110 and L40. It is worth mentioning that this result is not completely in agreement with the microarray analysis carried out by Funicane and coll. (2008) to profile gene expression changes in the bovine mammary gland at the onset of lactation using the Affymetrix Genechip Bovine Genome Arrays (23,000 gene probes) since these authors report that none of the milk protein genes showed significant changes in expression from late pregnancy (5 days before parturition) to early lactation (10 days after parturition). This is probably due (at least in part) to the larger gap existing between the last stage of pregnancy taken into account in our study (P110) which is *ca*. 40 days ahead of parturition, instead of 5 days in the Funicane's study. However, it is worth mentioned that in mice, the expression of milk protein genes increases progressively during pregnancy, especially in the last part of pregnancy, and that there is no sharp increase at secretory activation (parturition) except for δ-casein (Anderson *et al.*, 2007).

In the pregnant goat mammary gland, milk protein genes were not the only genes expressed. The expression of genes involved in lipid biosynthesis and secretion was also recorded. Indeed, the microarray analysis showed that adipophilin (*ADRP*), acyl CoA:diacylglycerol transferase 1 (*DGAT1*), fatty acid binding protein 4 (*FABP4*) and fatty acid synthase (*FASN*) are expressed in the goat mammary gland all along pregnancy. This result which was confirmed and validated in qPCR, is not really surprising since fatty acid synthase activity has been already measured in the pregnant mammary gland of cows (Mellenberger *et al.*, 1973) and rabbits (Mellenberger & Bauman, 1974) and *ADRP* mRNA was also found in pregnant mouse mammary gland (Anderson *et al.*, 2007). Such a result can be interpreted by the fact that during pregnancy, cell proliferation requires the synthesis of many membranous lipids composed of long chain fatty acids (as reviewed by Mellenberger *et al.*, 1973). Thus, the lipid synthesis activity observed in the goat mammary gland during pregnancy would not be dedicated to milk synthesis and secretion but more related to the cellular proliferation associated with mammary gland development which needs membrane biosynthesis. The expression of *DGAT1* gene during pregnancy contributes to epithelial proliferation and alveolar development of the mammary tissue. This is in agreement with the fact that mice

lacking the DGAT1 enzyme have impaired mammary gland development, characterized by decreased epithelial proliferation and alveolar development (Cases *et al.*, 2004).

Nevertheless, genes involved in milk lipid biosynthesis as *LPL*, the long chain acylCoA synthetase homolog 1 (*ACSL1*) and lipin1 (*LPIN1*), as well as in lactose synthesis, such as *UGP2* and *B4GALT1* (metacluster D and cluster 17) are only expressed in lactation in the goat mammary gland (Table II.5). Likewise, Akt1, a serine/threonine protein kinase which plays a key role in the regulation of glucose transport and lipid metabolism (Anderson *et al.*, 2007; Schwertfeger *et al.*, 2003) displays a significant increase in expression, between P110 and L40 (metacluster D). Butyrophilin (*BTN1A1*) which is supposed to be an activator of milk secretion (Ogg *et al.*, 2004) followed the same pattern (Table II.5 and Figure II.3A). *SREBF1* (Sterol regulatory element binding transcription factor 1) which represents a central node in the milk lipid metabolism network and controls transcription of most of the genes which regulate milk fat synthesis in mice (Rudolph *et al.*, 2003) and cattle (Bionaz & Loor, 2008), was also shown in our study to be a critical regulator of transcription as many genes regulated by this factor showed the same expression profile (*ACACA*, *FASN*, *LPL*, *ACSL1* and *INSIG1*).

In addition, α_{s1}-casein gene (*CSN1S1*), of which the expression profile is somewhat different from that observed with the other casein transcripts, as confirmed by qPCR experiments (Figure II.3, pattern B), seems to be significantly expressed only in lactation. This observation is in agreement with the notion that α_{s1}-casein could be necessary for caseins secretion, in the goat mammary tissue (Chanat *et al.*, 1999).

The comparison between the last pregnant developmental stage and lactation in our study and the study carried out by Finucane and coll. (2008) shows only 3 common genes in the two top 50 genes differentially expressed (lactoperoxidase, lipoprotein lipase and lipin 1). The Finucane et al. study stages compare 5 days before and 10 days after parturition, while our study compare 40 days before and after parturition. Thus, the comparison between these 2 studies helps to hypothesize that for the *LPO*, *LPL* and *LPIN1* genes the expression profile changes at the parturition and these genes are up-regulated in lactation. While, for the genes that show an increased expression in lactation only in our study (*CSN1S1*, *CSN1S2*, *LALBA*, *BTN1A1*, *FABP3*, *SCD*), given the results on mouse mammary tissue (Anderson *et al.*, 2007; Rudolph *et al.*, 2003; Rudolph *et al.*, 2007), we can hypothesize that these transcripts follow a progressive increase until parturition.

Related to that, general biological functions associated with lactating mammary gland validate the conclusion, supported by our study, that the lactating stage is characterized by molecular transport, cellular movement and protein trafficking. Thus, the activation of secretion seems to appear only after parturition in the mammary gland of ruminants as it is for cow mammary gland (Finucane *et al.*, 2008).

Expression of genes of immunity during pregnancy marks tissue remodeling and differentiation

This study describes the expression of immune cell signaling genes in a pregnant mammary gland of nulliparous ruminants. Genes with expression profile of IDID (increased expression at mid and 110 days of pregnancy) code for proteins of the innate and specific immunity as classified thereafter.

Genes of the innate immunity are coding for proteins of the complement system (*C1QB*, *CD2*, *CD3E*, *CD69*, *CD84*, *CD96*), interferons and interferon receptors (*IRF1*, *IRF2*, *IRF8*, *IFNAR1*), interleukins and interleukin receptors (*IL18*, *IL2RG*), chemokines and chemokine receptors (*CCL5*, *CXCL10*, *CXCR4*), receptor of natural killer cells (*KLRK1*) or macrophages (*MSR1*). We also recorded the expression of genes coding for receptors involved in the innate response activation to viral infection (*TLR3*) or protein that are the effectors of the innate immunity, such as lactoferrin and lyzosomal proteases (cathepsin C). Genes of specific immunity were mainly genes coding for immunoglobulin receptors with high affinity for Fc portion of IgG (*FCGR3A*) or IgE (*FCER1G*), or bovine MHC (*BoLA-A*, *BoLA-DMB*, *BoLA-DRB1*). There were also effectors of the immune response with the expression of genes encoding proteins of the multicatalytic proteinase complex immunoproteasome (*PSMB8*, *PSME1*, *PSME2*).

Several of the proteins quoted above are known to be chemoattractant or activator for blood monocytes, memory T helper cells and/or eosinophils. Indeed, IL18 acts on natural killer cell activity, lymphocyte T helper 1 (Th1) interferon gamma (IFN-γ) production and Th1 cell proliferation (reviewed in ref Suzuki *et al.*, 2003). CCL5 is a chemoattractant for blood monocytes, memory T helper cell and eosinophils (Proudfoot *et al.*, 1999). When binds to CXCR3, CXCL10 stimulates monocytes, natural killer and T-cell migration (Romagnani *et al.*, 2001).

The expression of monocytes attractant by the goat pregnant mammary gland suggests that there is a migration of macrophages and eosinophil cells from the blood to the mammary tissue. This phenomenon has been already reported in mouse mammary gland development

and was associated to matrix remodeling (reviewed in ref. Goudon-Evans *et al.*, 2002). Thus, the goat mammary gland differentiation could comprise two phases of tissue remodeling, one at mid and one in the last third of pregnancy (P70 and P110), characterized by the phagocytosis of one cell-type (adipocytes) or stroma, the development and the differentiation of another cell-type (epithelium). This hypothesis is confirmed by the up-regulation of genes coding for proteins working in the ruminant mammary gland immune system when facing mastitis exemplified by C1q and lactoferrin. C1q is essential in the classical pathway of the complement system (Rainard & Riollet, 2006), whereas lactoferrin has been reported to be mainly expressed by the mammary epithelial cells lining the ducts and the cistern (Molenaar *et al.*, 1996) and to function as a iron captor preventing bacterial development (Rainard & Riollet, 2006; Sordillo & Streicher, 2002). Other proteins are specific of epithelial cells such as TLR3 which is mainly described in intestinal epithelial cells of normal mucosa (Cario & Podolsky, 2000).

The notion according to which the goat mammary gland expresses a large number of innate immune genes in pregnancy, is consistent with the idea developed by Vorbach *et al.* (2006) that the mammary gland originates in the innate immune system. The innate immune system, also known as nonspecific or nonimmune responsiveness, is a first-line defense system that provides organisms with various immediately available defense mechanisms during the early stages of the defense system. Such a mechanism appears crucial for a tissue exposed to multiple etiological agents.

Moreover, clustering analysis shows a decreased expression in lactating goat mammary gland of the IL4-signalling pathway (metacluster E), characterized by the down-regulation of *IL4R* and *NFATc2* genes. Proteins encoded by these genes are lymphocyte T helper 2 (Th2) cytokines that participate to mouse mammary gland development *in vivo* (Khaled *et al.*, 2007). The secretion of these cytokines by mouse mammary epithelial cells (MEC) represents the induction of the differentiation process (Khaled *et al.*, 2007), and shows the role for immune cell signaling proteins in the MEC fate and function (Watson, 2008). The goat mammary gland expresses at 46 days of pregnancy, genes of the IL4-signalling pathway, this suggests that at this stage, the differentiation process would be already engaged. The decreased expression of the genes encoding these proteins in lactation is consistent with the notion that 40 days of lactation correspond to a well-established lactation stage, in which MEC are fully differentiated and that the mammary gland then does not express differentiation markers.

Taken together, all these data suggest an increased expression in the mid-pregnant goat mammary gland of a large number of genes encoding proteins of the immune system and milk proteins. We can hypothesize that a common regulator pathway exists for the expression of immunity and milk protein genes. The increase in expression of immune genes at mid-pregnancy underlines the crucial role of immune proteins in the remodeling step of the differentiation process.

II.1.5.Conclusion

Regarding general mechanisms underlying goat mammary gland development, we can conclude that during the first half of pregnancy the mammary tissue is switching from a stromal tissue to an epithelium. This notion is substantiated by the down-regulation of adipocyte specific genes (*e.g.* adiponectin or *ADIPOQ*) and the up-regulation of chemokine genes which are supposed to attract monocytes into the goat mammary gland and participate to its remodeling. The last half of pregnancy can be characterized by the proliferation of the mammary gland cells. As for numeral mammals (mice, rabbits, cows), secretion in the goat mammary gland is activated only after parturition. However, since the induction of the differentiation process seems to take place before 46 days of pregnancy, getting an insight into this mechanism would necessitate to screen earliest developmental stages during the first third of pregnancy.

Genes of immunity with increased expression during pregnancy participate to the remodeling step of the differentiation process. A subset of these genes might be activated to provide the neonate intestine and the mammary tissue of the lactating goat with non specific immunity. More than protective function, immune genes expressed in the mammary tissues participate to the mammary gland development and differentiation.

Here, we provide gene expression profiles of the whole mammary tissues. In ruminants, the mammary epithelium is developing in the stroma. Molecular cross-talks exist between these two histological structures. Therefore, it would be relevant to characterize precisely their relative contribution in the differentiation process. Powerful technologies, such as laser capture microdissection make it possible. This is currently under investigation in our laboratory.

II.1.6.Acknowledgments

We grateful acknowledge UCEA team for taking care of animals, Marie-Laure Martin-Magniette from AgroParisTech for her advice in the microarray experimental design, Philippe Bardoux and Cedric Cabau from SIGENAE team for their help in GO data publication and gene annotation, Karine Hugot and Diane Esquerre from CRB-GADIE for providing 22K bovine oligoarrays.

II.2. Comparaison directe 46 jours de gestation et 40 jours de lactation

La description morphologique du développement de la glande mammaire chez des génisses effectuée par (Swanson & Poffenbarger, 1979), laisse supposer que chez les ruminants et donc possiblement chez la chèvre la glande mammaire présenterait, en début de gestation, un tissu fibreux contenant des cellules épithéliales non différenciées, alors que la glande mammaire en lactation serait constitué de cellules épithéliales différenciées ayant développé un phénotype sécrétoire. Ces deux glandes apparaissent donc fonctionnellement très éloignées. Dans l'analyse globale des cinq stades de développement (décrite dans la partie précédente), il n'était donc pas possible de réaliser la comparaison directe entre le stade le plus précoce de la gestation (46 jours de gestation, G46) et le stade de lactation (40 jours, L40). En effet, dans le plan expérimental en boucle présenté dans l'article, si deux stades très différents sont comparés directement, cela induit un déséquilibre du plan expérimental et une diminution de sa puissance, en raison du poids trop important qu'aurait fait peser la comparaison G46-L40 sur l'ensemble du dispositif.

Dans cette étude, la comparaison directe entre ces deux stades, nous a permis de produire des listes de gènes exprimés en début de gestation et en lactation et de confirmer le plan expérimental de l'analyse précédente.

II.2.1.Matériels et méthodes

II.2.1.1. Préparation des cibles complexes

L'extraction des ARN totaux à partir de tissus mammaires a suivi la même procédure que celle décrite dans la partie « Matériels et Méthodes » du paragraphe II.1.

Les ARN totaux extraits ont ensuite été traités à la DNase en utilisant le kit SV Total RNA Isolation System (Promega, France). La quantité d'ARN a été évaluée par détermination du ratio 260/280 nm au moyen d'un spectrophotomètre de type nanodrop ND-1000 (Nanodrop Technology, Nixor Biotech, France). La qualité de chaque échantillon a été estimée avec le Bioanalyzer 2100 (Agilent Technologies, France), en utilisant la puce Nano 6000. Le Bioanalyzer permet d'estimer la qualité d'un échantillon en donnant un RIN ou « RNA Integrity Number », qui correspond, sur une échelle de 1 à 10, à une estimation du degré de d'intégrité de l'ARN. Les échantillons dont le RIN était inférieur à 7,4 étaient considérés comme étant de qualité insuffisante (en partie dégradés) et par conséquent écartés de l'expérience.

Les ARN totaux ont ensuite été rétro-transcrits en ADN complémentaire (ADNc) et marqués directement en utilisant le kit « Pronto ! Plus Direct System » (Promega, France), en suivant les instructions du fabricant. Brièvement, les ARN totaux ont été incubés pendant 10 min à 70 °C en présence d'amorces « random » et d'amorces oligo(dT) pour l'amorçage de la réaction de rétro-transcription. Les marquages s'effectuaient directement au cours de la réaction de transcription inverse, pendant 2h à 42°C en utilisant l'enzyme ChipShot™ Reverse Transcriptase et les nucléotides modifiés (Cy®3 dCTP et Cy®5 dCTP). La réaction de rétro-transcription était stoppée en incubant le mélange réactionnel à 37°C pendant 15 min en présence de RNase H. Cette enzyme est une hexo ou endonucléase spécifique des hétéroduplexes ARN:ADN. Elle hydrolyse la plus grande partie de l'ARN après sa transcription inverse en ADN.

Les ADNc marqués ont été séparés des cyanines non incorporées sur des colonnes ChipShot™ (Promega, France). La fréquence d'incorporation des fluorochromes (« Frequence of Incorporation » ou FOI) et la concentration d'ADNc (en pmol/µl) ont été estimées en mesurant les absorbances à 260 nm, 550 nm et 650 nm, au moyen du Nanodrop. Les ratios suivant ont été calculés : FOI (Cy3) = (((Abs550 x 60) / 0,15) x 324.5) / (Abs260 x 37 x 60), FOI (Cy5) = (((Abs650 x 60) / 0,25) x 324.5) / (Abs260 x 37 x 60). Lorsque les FOI étaient inférieures à 20 pour la cyanine 3 et 12 pour la cyanine 5, la qualité du marquage était jugée insuffisante et les échantillons étaient rejetés. Ces valeurs sont les seuils fixés par Promega pour garantir une qualité optimale de l'expérience de profilage expressionel sur microarrays.

II.2.1.2. Hybridations des cibles complexes sur les lames Operon

Les cibles complexes ont été hybridées à raison de 65 pmol de cible marquée, sur des lames UltraGaps (Corning B.V. Life Sciences) contenant le set de 8 329 oligonucléotides bovin commercialisé par Opéron auquel étaient ajoutés 50 oligonucléotides représentant pour l'essentiel des gènes impliqués dans la lipogenèse (numéro d'accession GEO :GPL4594 (Ollier *et al.*, 2007). Les oligonucléotides ont été déposés sur les lames en dupliquas (deux spots par séquence oligonucléotidique) sur la plateforme « Biopuces » de la Genopôle de Toulouse. Les 19 200 spots présents sur la lame correspondent à 8 379 gènes différents (sondes) dont 6 656 (soit 79,4 %) possédaient une annotation reconnue par le logiciel Ingenuity Pathway Analysis (IPA).

Les cibles ont été déposées sur les lames selon un plan expérimental comparant deux conditions, avec 3 répétitions biologiques effectuées en dye-swap (soit 6 lames au total). Avant l'hybridation, les lames ont été pré-hybridées en utilisant le kit Pronto ! Universal microarray Hybridation (Promega, France). L'hybridation a été effectuée au bain marie pendant 16h à 42°C dans des chambres à hybridation Corning (Corning Life Science, France). Le lavage des lames a été réalisé en utilisant les solutions et les instructions fournies avec le kit Pronto ! Universal microarray Hybridation (Promega, France). Le scan a été effectué sur le scanner Genepix 4 000B de la plateforme Puces à ADN de l'Institut Pasteur (Paris) à une résolution de 10 µm. L'intensité du signal pour chaque spot a été quantifiée sur le logiciel Imagene (Biodiscovery, USA).

II.2.1.3. Traitement des données issues de la quantification

Les fichiers quantifiés ont été normalisés en utilisant la fonction « *normalisation* » du package « anapuce » du logiciel R qui procède d'abord par une normalisation Lowess suivie d'une correction des signaux par bloc.

La cohérence du fichier des signaux normalisés a été vérifiée en utilisant la fonction « *verifdonnee* » développée sous R. Cette fonction calcule (i) les coefficients de corrélation entre les lames, pour vérifier que les swaps présentent des signaux opposés, (ii) les coefficients de corrélation entre les intensités des signaux normalisés de toutes les lames, pour vérifier que l'effet « lame » n'est pas plus fort que l'effet « échantillon » et que les répétitions biologiques d'une même condition sont bien corrélées. Cette fonction trace également des box-plot ce qui permet de visualiser rapidement si les lames sont comparables entre elles.

L'analyse différentielle a été faite en utilisant la fonction « *DiffAnalysis* » du package « anapuce » sous le logiciel R, en classant les gènes par groupe de variance homogène (Delmar *et al.*, 2005). La correction des probabilités critiques a été effectuée par la méthode de (Benjamini & Hochberg, 1995). Lorsqu'il existait plusieurs spots correspondant à la même séquence oligonucléotidique, la moyenne des intensités du signal a été calculée.

Le niveau d'expression de chaque gène a été évalué pour chaque stade de développement en calculant l'intensité du signal et en considérant qu'un gène est (i) fortement exprimé lorsque son intensité est supérieure à 10% de l'intensité la plus élevée de l'ensemble des données, (ii) faiblement exprimé lorsque son intensité est inférieure à dix fois l'intensité la plus faible, (iii) moyennement exprimé lorsque son intensité est comprise entre les deux critères précédents.

II.2.1.4. Validation des différentiels d'expression sur une vingtaine de gènes par PCRq

Les différentiels d'expression ont été validés pour certains gènes par la quantification de leur transcrits en PCRq. Les systèmes PCR analysés et la méthodologie étaient les mêmes que pour la partie II.1.

Pour comparer les résultats obtenus par PCR avec ceux obtenus par la technique des « microarrays », chaque donnée normalisée de la quantification par PCRq et du signal de fluorescence correspondant sur « microarrays » a été calibrée dans un premier temps relativement à la valeur correspondante des chèvres en lactation. Ensuite la moyenne et l'écart-type à 46 jours de gestation et 40 jours de lactation ont été calculées et des histogrammes ont été tracés.

II.2.2.Résultats

II.2.2.1. Analyse des puces à ADN

Afin de déterminer les différences d'expression des gènes entre la glande mammaire en début de la gestation (G46) et celle en lactation (L40), une analyse différentielle a été effectuée à partir des données issues de la comparaison directe des ARN extraits de ces deux types de tissu, sur « microarrays » en dye-swap. L'analyse différentielle a été effectuée à partir de trois répliquas biologiques, répétés techniquement deux fois en Cy3 et Cy5 (dye-swap).

En classant les gènes par groupe de variance homogène et en fixant le seuil de significativité à 1%, 575 sondes (soit 7% des sondes présentes sur la lame) ont été déclarées différemment exprimées entre 46 jours de gestation et 40 jours de lactation. Ces 575 sondes représentaient 502 gènes différents avec une annotation acceptée par le logiciel d'interprétation biologique, Ingenuity Pathway Analysis (IPA).

Une seule sonde présentait un fort signal aux deux stades physiologiques considérés. Le gène correspondant spécifie la sous-unité 3 du complexe intégrateur (spliceosome) participant à l'activité des petits ARN nucléaires (snRNA) qui interviennent dans les mécanismes d'épissage des ARN. Trois sondes étaient très fortement exprimées en lactation et faiblement exprimées en début de gestation. Ces sondes correspondent au précurseur de l'α-lactalbumine et aux gènes spécifiant les caséines α_{s1} et κ. Les autres sondes présentaient un signal moyen ou faible.

Parmi les sondes déclarées différemment exprimées, 295 étaient surexprimées en lactation et 280 l'étaient à 46 jours de gestation. Si le seuil de différence d'expression était fixé à deux, c'est à dire que l'expression d'un gène dans une condition est deux fois supérieure à son expression dans l'autre condition, alors 112 sondes étaient deux fois plus exprimées en lactation qu'en gestation et 28 sondes étaient deux fois moins exprimées en lactation qu'en gestation.

De façon attendue, les gènes surexprimés en lactation étaient représentatifs des fonctions biologiques correspondant à la synthèse des protéines et des lipides (cf. Tableau II.7).

Tableau II-7 : Liste des gènes différemment exprimés au niveau du tissu mammaire entre 46 jours de gestation et 40 jours de lactation. Le seuil de significativité est de 1% avec la correction des probabilités par la méthode de Benjamini-Hochberg.

	Nom du gène	Code du gène	Ln2(R/V)	Ratio d'expression
Gènes surexprimés en lactation				
Métabolisme des protéines				
Synthèse des protéines	Caséine kappa	*CSN3*	7,4	165,4
	Alpha-lactalbumine	*LALBA*	7,3	158,6
	Caséine alpha S1	*CSN1S1*	6,3	76,3
	Caséine beta	*CSN2*	4,7	26,4
	Beta-lactoglobuline	*LGB*	4,3	20,0
	Polymerase alpha 2	*POLA2*	1,0	2,0
Modifications post-traductionnelles	DNAJ	*DNAJB11*	4,5	22,7
		DNAJC12	1,5	2,7
	Defective in cullin neddylation	*DCN1*	2,3	5,0
	WD repeat domain 34	*WDR34*	2,0	4,1
	Ubiquitin specific peptidase 46	*USP46*	2,0	4,1
	Ubiquitin specific peptidase 9	*USP9Y*	1,0	2,0
Mouvement de protéines	Recepteur de la particule de reconnaissance du signal	*SRPR*	1,6	3,0
	Trafficking protein complex	*TRAPPC4*	1,6	3,1
Synthèse des lipides	Leptine	*LEP*	3,3	9,7
	Stearoyl-CoA desaturase	*SCD*	2,7	6,5
	Leptine receptor	*LEPR*	2,3	5,1
	Phosphatidylethanolamine-N-methyltransferase		2,0	4,1
	Lipoproteine lipase	*LPL*	2,0	4,0
	Fatty Acid Binding Protein 3	*FABP3*	1,9	3,7
	Lipidosine (Acyl-CoA synthase)	*ACSBG1*	1,7	3,2
	Glycerol-3-phosphate acyltransferase	*GPAM*	1,4	2,7
	Malate dehydrogenase 2	*MDH2*	1,2	2,3
Composants membrane du globule gras	Butyrophiline	*BTN1A1*	3,0	7,8
	Mucine 15	*MUC15*	1,5	2,9
	Milk fat globule-EGF factor 8	*MFGE8*	1,5	2,8
Gènes surexprimés en gestation				
Mouvement cellulaire	Flovoprotéine transfert d'électron	*ETFB*	-2,6	6,3
	Lactophiline 1	*LPHN1*	-1,2	2,3
	Synucléine alpha	*SNCA*	-1,1	2,2
Cycle cellulaire	Protéine d'intéraction à la Thiorédoxine	*TXNIP*	-1,7	3,3
	Sestrine 1	*SESN1*	-1,2	2,3
Organisation et assemblage cellulaire	Protéine de liaison à la guanine	*GNB2L1*	-1,0	2,0
	Collagène type III alpha 1	*COL3A1*	-2,4	5,2
	Protéine de liaison au ribosome	*RRBP1*	-2,1	4,4
	Protéine de liaison au TGF beta	*LTBP3*	-1,3	2,5
	Thymosine	*TMSL3*	-1,3	2,4
	Collagène type VI, alpha 2	*COL6A2*	-1,2	2,3
	Domaine associé à la fas	*FADD*	-1,1	2,1

La fonction « synthèse protéique » était caractérisée par des gènes spécifiant (i) les protéines du lait telles que l'α-lactalbumine, la β-lactoglobuline et les caséines κ, $\alpha_{s1,}$ β et à un seuil de significativité supérieur la caséine α_{s2} (P=0,012) ; (ii) les protéines qui interviennent dans le repliement et les modifications post-traductionnelles des protéines telles que *DNAJ*, *DCN1* (« Defective in cullin neddylation », la nédyllation étant une modification post traductionnelle des protéines), le domaine répété WD 34 (*WDR34*, facilitant la formation de complexes multiprotéiques) et les peptidases spécifique de l'ubiquitine (« Ubiquitin specific

peptidase », *USP46, USP9Y*, intervenant dans la dégradation de protéines) ; (iii) les protéines responsables du transport inta-cellulaire des protéines telles que le récepteur de la particule de reconnaissance du peptide signal (« Signal Recognition Particle Receptor », *SRPR*) et le « trafficking protein complex » (*TRAPPC4*) qui facilitent l'entrée des protéines dans la voie de sécrétion, et assurent leur orientation vers le réticulum endoplasmique.

La fonction « biosynthèse lipidique » se caractérise par les gènes spécifiant les protéines qui contribuent ou interviennent directement dans la synthèse des lipides tels que la leptine et son récepteur (*LEP*, *LEPR*), la stéaroyl-CoA désaturase (*SCD*), la phosphatidyléthanolamine-N-methyltransférase, la lipoprotéine lipase (*LPL*), la « fatty acid binding protein » (*FABP3*), la lipidosine (*ACSBG1*), la Glycerol-3-phosphate acyltransférase (*GPAM*) et la malate déhydrogénase (*MDH2*) et par les gènes spécifiant les protéines impliquées dans la sécrétion du globule gras telles que la butyrophiline (*BTN1A1*), la mucine 15 (*MUC15*) et le « milk fat globule-EGF factor 8 » (*MFGE8*).

Les gènes surexprimés en gestation étaient représentatifs des fonctions biologiques, définies par IPA, telles : le « mouvement cellulaire », « cycle cellulaire » et « organisation et assemblage cellulaire ».

Les gènes caractérisant la fonction « mouvement cellulaire » spécifient : la sous unité β d'une flavoprotéine qui joue le rôle d'accepteur d'électrons de déshydrogénases (« Electron Transfer Flavoprotein », *ETFB*) et intervient également dans le catabolisme de certains acides gras, la latrophiline 1 (*LPHN1*) qui appartient à une sous-famille des récepteurs couplés à des protéines G, dotées de fonctions de signalisation mais aussi d'adhésion et enfin la synucléine α (*SNCA*), une petite protéine (14 kD) non repliée, possédant les caractéristiques d'une chaperonne (participe à la formation de complexes SNARE dont la fonction première est de médiatise l'ancrage des vésicules sécrétoires aux « porosomes » membranaires des cellules eucaryotes) principalement exprimée dans le cerveau. Cette protéine qui est aujourd'hui l'objet d'une attention particulière en raison de son implication dans des désordres neurodégénératifs tels la maladie de Parkinson (Emadi *et al.*, 2009), jouerait également un rôle clef dans le trafic cellulaire.

Les gènes impliqués dans le cycle cellulaire spécifient un régulateur de transcription : la protéine d'interaction à la Thiorédoxine (*TXNIP*), et la sestrine 1 (*SESN1*) qui régule négativement la prolifération cellulaire.

Les gènes participant à l'organisation et l'assemblage des cellules spécifiaient les protéines du collagène (*COL3A1* et *COL6A2*), glycoprotéine sécrétée par le tissu conjonctif,

la thymosine (*TMSL3*) qui intervient dans l'organisation du cytosquelette et le domaine associé à la fas (*FADD*), un activateur de l'apoptose.

Devant le faible nombre de gènes surexprimés en gestation (n=28), nous avons listé les gènes pour lesquels l'intensité du signal en début de gestation était non négligeable (supérieure à 10 fois l'intensité la plus faible de l'ensemble des données). Les gènes remarquables de cette liste sont ceux spécifiant les protéines du cytosquelette telles que les actines (*ACTG1*, *ACTN2*), les myosines (*MYH3*, *MYH7*, *MYH9*, *MYH10*, *MYH14*, *MYL3*, *MYL4*, *MYO10*, *MYRIP*, *MYO1A*) et les kératine (*KRT6A*, *KRT7*, *KRT14*, *KRT17* et *KRT20*) ; et certains gènes des lactoprotéines (*CSN2* spécifiant la caséine κ et *BLG* spécifiant la β-lactoglobuline).

II.2.2.2. Validation des données de microarray par PCR quantitative en temps réel

Une vingtaine de gènes spécifiques de la biosynthèse des lactoprotéines, de la lipogenèse ou du développement de la glande mammaire ont été quantifiés par PCRq. Ces données ont ensuite été confrontées aux résultats obtenus par la technique des « microarrays » (Figure II.6), avec les mêmes échantillons biologiques.

La confrontation des résultats entre les deux méthodes de quantification des transcrits rend compte de quelques aberrations. Ainsi, le nombre de transcrits du gène spécifiant l'Acétyl CoA Carboxylase (*ACC*) est surestimé pour la lactation avec la PCRq, s'expliquant par une valeur exagérée pour une chèvre lactante. Les autres résultats mettant en évidence un biais entre « microarray » et PCRq s'observent pour les données d'expression obtenues sur les gènes *ADRP*, *SCD*, *DGAT1*, *MFGE8*, *GlyCAM1*, *LPO* et *Elf5*. L'intensité du signal de la sonde de ces gènes est moyenne ou forte à la fois en gestation et en lactation sauf pour la sonde spécifiant le gène *SCD*, où le signal est faible sur toutes les lames. Pour cette sonde, la séquence oligonucléotidique a été comparée avec le génome bovin et elle ne matche pas avec la séquence du gène de la SCD, cette sonde ne caractérise plus la SCD. Les séquences oligonucléotidiques des autres sondes ont également été vérifiées sur le NCBI, aucun problème n'a été relevé. De plus, il n'y a pas de saturation pour ces sondes sur les lames. Il n'y a donc pas d'explication triviale à l'observation d'une expression non négligeable de ces gènes en début de gestation alors qu'il y a peu de transcrits dénombrés en PCRq. Toutefois, pour ces gènes la tendance suivie entre les deux méthodes d'analyse d'expression est

similaire. Le résultat aberrant pour les données de microrray obtenues pour la caséine α_{s2} s'explique par une intensité du signal proche du bruit de fond sur toutes les lames sauf pour deux, correspondant à une chèvre en lactation. Il se pourrait que la séquence oligonucléotidique de la sonde correspondant à la *CSN1S2* ait été déterminée dans une région d'epissage alternatif, représentatif d'un variant de cette caséine, et que les chèvres analysées ici ne soient pas de même génotype pour ce gène.

Certains gènes de la lipogenèse, tels que les gènes spécifiant l'adipophiline (*ADFP*), la *FABP4*, *DGAT1*, la « fatty acid synthase » (*FAS*) et la Stéaroyl CoA Desaturase (*SCD*) sont à la fois exprimés en gestation et en lactation avec un nombre de transcrits plus important en lactation (de 2,5 à 10 fois plus en se basant sur les quantifications de PCRq). Les autres gènes quantifiés en PCRq spécifiant des protéines qui participent à la synthèse et à la sécrétion du globule gras tels que GlyCAM 1, la butyrophiline, MFGE8 ne sont exprimés qu'en lactation. De même, tous les gènes spécifiant les protéines du lait ne sont exprimés qu'en lactation. Certains gènes spécifiant les protéines qui interviennent dans le développement de la glande mammaire, tels que l'ostéopontine, Id2 et la kératine 14 (protéine du cytosquelette, spécifique des cellules myoépithéliales) sont exprimés à la fois en gestation et en lactation de façon équivalente. Par contre, les gènes spécifiant Elf5 et la lactopéroxydase seraient quasi exclusivement exprimés en lactation.

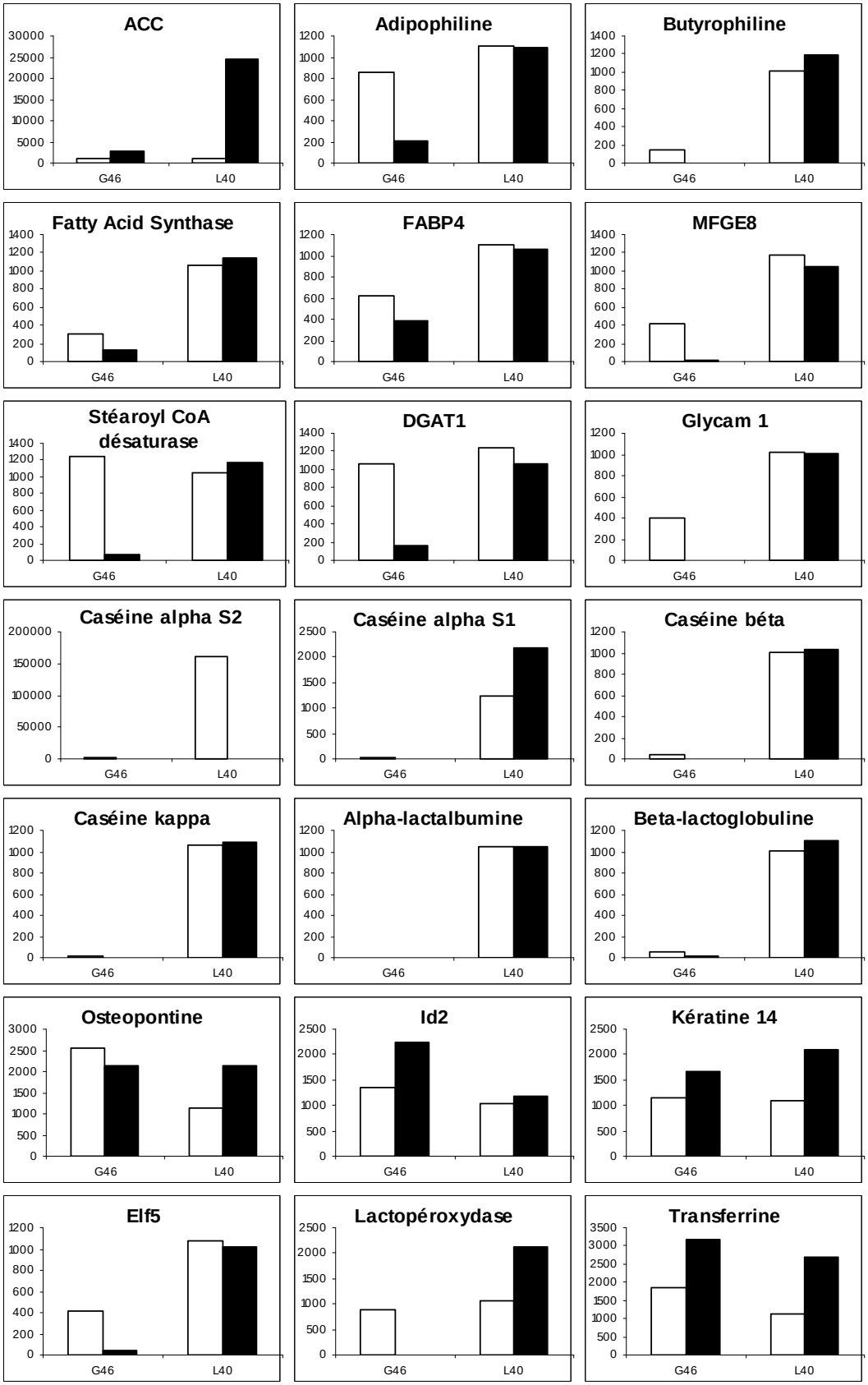


Figure II-6 : Comparaison des profils d'expression d'une vingtaine de gènes impliqués dans la synthèse et la sécrétion des lipides et protéines du lait ou impliqués dans la mise en place du tissu mammaire, obtenus par la technique d'analyse sur microarrays (barres blanches) et en PCRq (barres noires).

II.2.3.Discussion

II.2.3.1. La glande mammaire lactante : dominance d'un épithélium sécrétoire actif

Les tissus mammaires caprins en lactation sur-expriment les gènes spécifiant les lactoprotéines telles que les caséines, l'α-lactalbumine et la β-lactoglobuline. Ces gènes sont jusqu'à 150 fois plus exprimés en lactation qu'en gestation. Or, ces gènes sont spécifiques de la glande mammaire en lactation puisque ces protéines caractérisent le lait (Tremblay *et al.*, 2003). Des études sur les tissus mammaires murins (Anderson *et al.*, 2007; Robinson *et al.*, 1995; Rudolph *et al.*, 2003) et de rat (Nakhasi & Qasba, 1979) ont montré que les gènes spécifiant les protéines du lait n'étaient pas exprimés en début de gestation mais fortement exprimés en lactation, bien que des traces de caséine α_{s1} et de caséine β aient été détectées dans les tissus mammaires murins en début de gestation (Kanazawa & Kohmoto, 2002) et que des traces de caséines et de leur ARNm l'aient également été dans les tissus mammaires de lapine, en début de gestation (Shuster *et al.*, 1976).

Les gènes associés à la maturation et à la sécrétion des protéines sont également surexprimés en lactation. Ces gènes ne sont pas spécifiques des tissus mammaires lactants mais caractérisent un tissu sécrétoire. Le récepteur de la particule de reconnaissance du peptide signal (*SRPR*) et le « trafficking protein particle complex 4 » (*TRAPPC4*) sont des molécules qui participent au passage de la protéine nouvellement traduite du cytoplasme vers le réticulum endoplasmique, mécanisme propre aux protéines, telles que les lactoprotéines, qui sont dirigées vers la voie de sécrétion (Mepham *et al.*, 1992), et au transfert du RE vers le cis-Golgi, en prenant part, en amont de l'assemblage des « SNARE complex » à l'accrétion et à la fusion des vésicules sécrétoires.

Les gènes surexprimés dans les tissus mammaires caprins en lactation et spécifiant les protéines impliquées dans le repliement et les modifications post-traductionnelles des protéines (*DNAJB11, DNAJC12, DCN1, WDR34, USP46, USP9Y*) ne sont pas spécifiques des protéines du lait. En effet, les protéines du lait subissent essentiellement des phosphorylations et des glycosylations qui font intervenir un certain nombre d'enzymes au rang desquelles figurent des kinases et des transférases (Mepham *et al.*, 1992).

La glande mammaire caprine en lactation sur-exprime également des gènes spécifiques ou impliqués dans la synthèse et de la sécrétion des lipides du lait. En effet, les gènes spécifiant la leptine (*LEP*), la stéaroyl-CoA désaturase (*SCD*), la lipoprotéine lipase

(*LPL*), la « fatty acid binding protein » 3 (*FABP3*), la glycérol-3-phosphate acyltransférase (*GPAM*), la lipidosine (*ACSBG1*) et la butyrophiline (*BTN1A1*) sont spécifiquement exprimés dans les tissus de la glande mammaire en lactation (Bernard *et al.*, 2008; Bionaz & Loor, 2008).

D'autres gènes qui ne sont pas directement impliqués dans la synthèse des lipides du lait, mais participant à la synthèse des lipides dans d'autres tissus, sont également surexprimés en lactation. Ceci est le cas pour la phosphatidyléthanolamine-N-méthyltransférase qui intervient dans la synthèse des lipides membranaires (Bell *et al.*, 1981), la malate déhydrogénase, enzyme du cycle de l'acide citrique qui permet la production de NADPH utilisé dans la synthèse des acides gras (Hawke & Taylor, 1995) et la mucine 15 présente dans la membrane du globule gras et dont le rôle dans la sécrétion de celui-ci est encore hypothétique (Heid & Keenan, 2005).

II.2.3.2. La glande mammaire en gestation : dominance d'un tissu conjonctif

Devant le faible nombre de gènes significativement (2 fois) moins exprimés entre la lactation et la gestation (28 gènes), l'interprétation biologique est assez difficile. Toutefois, les gènes spécifiant le collagène sont largement surexprimés en gestation. Or, le collagène est un constituant majoritaire du tissu conjonctif de la matrice extracellulaire. L'expression plus importante des gènes spécifiant ces glycoprotéines en début de gestation dans les tissus mammaires caprins est en accord avec l'observation faite par (Rudolph *et al.*, 2003) sur les tissus mammaires murins. En début de gestation, les tissus mammaires seraient essentiellement composés de tissu conjonctif exprimant des gènes du maintien de l'activité de base d'une cellule, expliquant ainsi le faible nombre de gènes sous-exprimés en lactation.

Le niveau d'expression, non négligeable (bien supérieur au bruit de fond), en début de gestation de gènes des protéines caractéristiques de tissu musculaire (myosine, actine et cytokératine) ou de CEM (*CSN2* et *BLG*), suggère qu'à 46 jours de gestation les tissus mammaires caprins contiennent, en plus du tissu conjonctif, des cellules représentatives de tissus musculaires (probablement des cellules myoépithéliales) et de CEM. Ces gènes ne sont pas révélés par l'analyse différentielle puisque en lactation, les tissus mammaires contiennent en plus grande quantité les mêmes types cellulaires, l'expression de leurs marqueurs est donc plus importante à ce stade physiologique. Comme cela a été suggéré dans le premier article

(cf. Partie II.1), le déclenchement de la différenciation terminale de la glande mammaire caprine aurait lieu en tout début de gestation.

II.2.4.Conclusion

L'expérience de profilage expressionnel décrite ci-dessus était la première expérience de ce type réalisée dans le cadre de ma thèse. Elle met en évidence certains aléas de cette technique tels que l'incidence du bruit de fond, les défauts inhérents à la fabrication des supports et au « design » des oligonucléotides. Dans cette étude, les 21 gènes validés par PCRq ont permis de mettre en avant ces aléas et de noter qu'il est fortement recommandé, avec ce type de technologie, de vérifier *a posteriori* les gènes déclarés différemment exprimés et dans la mesure du possible, d'avoir une idée *a priori* sur la nature des gènes attendus, pour pouvoir contrôler spécifiquement leur profil d'expression.

Comme dans cette étude, le seuil de décision pour déclarer si un gène est différemment exprimé était très restrictif (1% avec la correction de Benjamini-Hochberg), vu le nombre relativement limité de sondes sur la lame (8 979), le fait que 575 sondes ont été déclarées différemment exprimées montre que la glande mammaire caprin en début de gestation est très différente de celle en lactation. La décision de ne pas inclure cette comparaison dans le plan expérimental de l'expérience décrite dans la partie II.1, était donc judicieuse.

Pour conclure, nos résultats de comparaison indiquent clairement que les tissus mammaires caprins en début de gestation possèdent les caractères d'un tissu conjonctif alors que les tissus mammaires en lactation présentent les caractéristiques d'un tissu sécrétoire exprimant les gènes de synthèse et de sécrétion des protéines et des lipides du lait, dont un répertoire a pu être dégagé à l'issue de cette étude. Ceci est cohérent avec les observations anatomiques faites sur la glande mammaire bovine en première gestation, par (Swanson & Poffenbarger, 1979) dans lesquelles il est décrit que les tissus mammaires en début de gestation ne contiennent pas de structure alvéolaire ni de produit de sécrétion. Toutefois, l'étude du niveau d'expression de certains gènes à 46 jours de gestation, indique, qu'à ce stade, les tissus mammaires sont plus complexes qu'un tissu conjonctif puisqu'ils expriment à un niveau non négligeable des gènes représentatifs de cellules myoépithéliales et de CEM.

Dans cette étude, l'utilisation de microarrays pour mesurer simultanément l'abondance de transcrits de milliers de gènes a été exploitée pour mettre en évidence une composition cellulaire des tissus non révélée par une analyse comparative.

III. Chapitre 3 : effet de la mutation K232A au locus *DGAT1* sur les profils d'expression génique des tissus mammaires bovin

Cette partie du travail de thèse s'inscrit dans le cadre d'un programme national : « Genomilk Fat », financé par l'ANR Genanimal et la société Apis-Gène. Ce projet a été initié en 2003, dans un contexte où le lait et les produits dérivés sont de plus en plus décriés par les consommateurs en raison de la forte teneur en acides gras saturés de la matière grasse laitière. Les acides gras saturés sont suspectés provoquer des maladies cardio-vasculaires lorsqu'ils sont consommés de façon excessive (Williams, 2000). Il est clairement établi que la composition du lait (teneur en matières grasse et protéique) peut être modifiée efficacement au moyen de leviers tels que l'alimentation animal (Chilliard *et al.*, 2000) et la Génétique (Grosclaude *et al.*, 1994; Khatkar *et al.*, 2004). Le programme « Genomilk Fat » avait ainsi pour objectifs d'étudier les effets de polymorphismes génétiques (loci *CSN1S1* chez la chèvre et *DGAT1* chez la vache), de l'alimentation et de l'interaction entre ces deux facteurs sur les performances laitières, la composition et les caractéristiques des structures supramoléculaires du lait et les profils d'expression génique du tissu mammaire. C'est sur la partie bovine de ce programme que le travail présenté ci-après a porté, en se basant sur le dispositif de détection de QTL laitiers mis en place sur le domaine expérimental du Pin-au-Haras (Inra, UE 326, Orne, France).

III.1. Facteurs de variation du taux butyreux et de la composition en acides gras du lait des ruminants

III.1.1. Les facteurs environnementaux modulant la matière grasse laitière

Le taux butyreux (TB) et la composition en acides gras (AG) des laits varient selon différents facteurs tels que la race, le potentiel génétique, la parité, le nombre de jours de lactation, l'état sanitaire et l'alimentation de l'animal ou d'autres facteurs attribués à l'environnement de l'animal, parfois moins maîtrisables, tel que la photopériode.

Le TB augmente en début de lactation, puis diminue au cours de la lactation.

La teneur en acide butyrique (C4:0) atteint un maximum dans le premier mois de la lactation puis diminue pour atteindre un minimum en fin de lactation (Parodi, 1974). Les teneurs en AG à chaîne courte, moyenne et saturée (acide caproïque, C6:0 à acide myristique, C14:0) augmentent pendant les deux premiers mois de la lactation, restent constantes jusqu'au cinquième ou sixième mois puis diminuent jusqu'à la fin de la lactation (Decean & Adda, 1970; Parodi, 1974). L'acide palmitique (C16:0) varie peu en cours de lactation (Decean & Adda, 1970; Parodi, 1974). Les teneurs en acides stéarique (C18:0) et oléique (C18:1) sont élevées en début de lactation, diminuent à mi-lactation et augmentent à nouveau en fin de lactation (Decean & Adda, 1970; Parodi, 1974). Le nombre de jours de lactation a peu d'effet sur la teneur en acides conjugués de l'acide linoléique (« conjugated linoleic acid », CLA) (Kelsey *et al.*, 2003). De même, la parité n'a pas d'effet sur la teneur en CLA (Kelsey *et al.*, 2003) ni sur la taille et la composition du globule gras du lait (pour revue Couvreur & Hurtaud, 2007).

L'alimentation constitue un moyen rapide et réversible pour modifier la teneur en matière grasse et la composition en AG des laits de ruminants. Les modifications sont dépendantes de la nature du fourrage, de son mode de conservation, du ratio fourrage/concentré et de la supplémentation des rations en lipides (Chilliard & Ferlay, 2004). Ainsi, par comparaison avec l'ensilage de maïs, le pâturage entraîne une diminution du TB et une augmentation de la production laitière (Chilliard *et al.*, 2000; Rego *et al.*, 2004). Une alimentation à base de foin ou d'herbe, dont la composition en AG est caractérisée par une forte teneur en acide linolénique (40%) augmente la teneur des laits en AG mono et polyinsaturés (Couvreur *et al.*, 2003; Hurtaud *et al.*, 2002; Rego *et al.*, 2004; Sauvant *et al.*, 2002).

Sur un régime à base d'ensilage de maïs, une supplémentation par des graines riches en acide linoléique (C18:2) ou linolénique (C18:3) (graines de tournesol ou de lin) provoquera une diminution de la production laitière et du TB alors que des graines riches en acide oléique (graines de colza) ne modifient pas significativement la production laitière (Casper *et al.*, 1988; Chilliard & Ferlay, 2004; Collomb *et al.*, 2004; Sauvant *et al.*, 2002). La graine de tournesol (riche en acide linoléique) entraîne une diminution plus importante de la teneur en AG saturés dans le lait et une augmentation plus forte de celle en AG insaturés que les graines de colza (riches en acide oléique), de lin (riche en acide linolénique) ou soja (riche en acide linoléique) (Casper *et al.*, 1988; Chilliard & Ferlay, 2004; Palmquist & Conrad, 1978; Sauvant *et al.*, 2002; Ward *et al.*, 2002).

La diminution de la teneur en AG saturés et l'augmentation de la teneur en AG insaturés dépendent également de la dose et de la forme d'apport des supplémentations lipidiques. Ainsi, la graine de lin crue diminue moins le rapport AG saturés sur AG insaturés que la graine de lin extrudée ou traitée par micro-ondes (Gonthier *et al.*, 2005). Il en va de même entre la graine de lin broyée et la graine de lin extrudée (Hurtaud *et al.*, 2006). L'apport croissant de lin dans la ration entraîne une augmentation de la production laitière et de la teneur du lait en acide linolénique ainsi qu'une diminution du TB et de la teneur en AG saturés (Hurtaud *et al.*, 2006).

III.1.2. Les polymorphismes génétiques impactant la composition des laits bovins

La notion de polymorphisme génétique s'applique lorsque l'on observe à un locus donné (c'est à dire un emplacement précis d'un chromosome) au moins deux allèles différents pour une même espèce. Le balisage du génome bovin par différents marqueurs (cartographie du génome bovin, pour revue Leveziel & Cribiu, 2000) a permis l'identification de nombreuses régions d'intérêt impliquées dans l'expression d'un caractère quantitatif (ou « Quantitative Trait Loci », QTL) affectant la composition générale et fine du lait, la résistance aux maladies et la reproduction. L'identification des gènes contenus dans une région d'intérêt peut permettre de connaître la ou les mutation(s) causales expliquant la variabilité du caractère associé à cette région. Le séquençage du génome bovin a permis une avancée importante dans l'identification des mutations causales associées aux QTL. Il est aujourd'hui possible, au moyen de puces dédiées, de réaliser un génotypage des animaux simultanément sur plus de 60 000 marqueurs moléculaires correspondant à des mutations ponctuelles, réparties sur l'ensemble du génome ou « single nucleotide polymorphism » (SNP).

Chez les bovins, les QTL majeurs affectant la production laitière ont été localisés sur les chromosomes bovins (BTA) 1, 3, 6, 9, 10, 14, 20 et 26. Les QTL caractérisant le TB ont été positionnés sur BTA6 et BTA14 et ceux expliquant la quantité de matière grasse laitière sont localisés sur les BTA14 et BTA26. Les QTL affectant le taux protéique (TP) sont localisés sur les BTA6, BTA20 et BTA26 et ceux expliquant la quantité de matière protéique sur les BTA6 et BTA26 (pour revue Jiang *et al.*, 2005 ; Khatkar *et al.*, 2004).

Ces approches cartographiques et analyses de liaisons avec des marqueurs polymorphes n'ont pas encore permis d'identifier les chromosomes et/ou les régions

chromosomiques impactant la composition fine en AG des laits, bien que les teneurs en AG saturés et mono-insaturés des laits soient des caractères héritables (Soyeurt *et al.*, 2007).

Dans la plupart des cas, les gènes contenus dans la zone (ou fragment) chromosomique correspondant du QTL, ne sont pas connus et leur nombre peut être relativement important. Dans certains cas cependant, on a pu identifier le gène et la mutation ayant un effet sur le caractère analysé. Pour les QTL affectant le TB ou la quantité de matière grasse, les QTL respectivement repérés sur les chromosomes BTA6, BTA14 et BTA26 ont été associés : à une mutation non sens au niveau du gène *ABCG2* (Cohen-Zinder *et al.*, 2005; Olsen *et al.*, 2007), à une double mutation dans l'exon 8 du gène *DGAT1* (Grisart *et al.*, 2002; Winter *et al.*, 2002) et à une mutation dans le cinquième exon du gène *SCD* (Taniguchi *et al.*, 2004).

III.2. Du QTL bovin affectant la quantité de matière grasse laitière à l'enzyme diacylglycérol acyltransferase 1 (DGAT1)

III.2.1. Du QTL affectant la quantité de matière grasse laitière au gène DGAT1

Le balisage du génome bovin par des marqueurs microsatellites a permis la détection de nombreuses régions d'intérêt ou QTL. Par la méthode de recherche de QTL en schéma « petites filles », Coppieters *et al.*, (1998) découvrent, à partir d'une population de vaches laitières de race Holstein-Friesian, un QTL affectant la production et la composition du lait (TB essentiellement) dans la terminaison centromérique du chromosome bovin 14 (BTA14). L'existence de ce QTL a ensuite été confirmée par (Heyen *et al.*, 1999) sur une population Holstein-Friesian d'Amérique du Nord et sur des Holstein israéliennes, en recentrant la position de ce QTL autour du marqueur D14S31. Riquet *et al.*, toujours en 1999, densifient la carte de la terminaison centromérique du BTA14 en positionnant 25 marqueurs microsatellites supplémentaires, ce qui leur permet de localiser ce QTL sur une région de 4,5 à 5 cM, entre les marqueurs ILST039 et BULGE004. Looft *et al.*, (2001), sur des populations de races Holstein allemande, Simmental allemande et Suisse Brune, associent le QTL au marqueur KIEL_08 et déterminent deux allèles notés 1 et 2, le génotype 2/2 étant caractérisé par une augmentation de production laitière (PL), de matière protéique sécrétée (MP) et une baisse de matière grasse sécrétée (MG) et de TB. Farnir *et al.*, (2002) reprécisent la zone de ce QTL dans la région proximale du marqueur ILSTS39.

En 2002, une analyse de génomique comparative entre le bovin, l'homme et la souris permet d'associer ce QTL bovin à une double mutation (GC→ AA) qui affecte deux

nucléotides successifs (10433-10434) de la séquence du gène (N° accession AJ318490 Genebank) spécifiant l'enzyme diacylglycérol acyltransferase 1 (DGAT1) (Grisart *et al.*, 2002; Winter *et al.*, 2002). Le gène *DGAT1*, localisé dans la terminaison centromérique de BTA14, est alors présenté comme un bon gène candidat pour expliquer le QTL dans la mesure où ce gène spécifie une enzyme qui catalyse l'étape ultime et limitante de la synthèse des triglycérides dans les hépatocytes (Mayorek *et al.*, 1989) et que l'invalidation de ce gène entraîne, chez la souris, l'absence totale de lactation (Smith *et al.*, 2000). La double mutation observée dans l'exon 8 du gène *DGAT1* bovin se traduit par la substitution d'un résidu lysine par un résidu alanine (K → A) au niveau de l'acide aminé 232 de la protéine. Ce polymorphisme est identifié comme la mutation K232A. Les dix-sept autres mutations observées sur le gène *DGAT1* bovin sont des mutations affectant un seul nucléotide (SNP). Elles se trouvent soit dans des séquences introniques, soit dans la région 5' non traduite (5' « UnTranslated Region », UTR) (Winter *et al.*, 2002) et n'ont pas d'effet reconnu sur les performances laitières. L'allèle codant pour le variant K232 semble être l'allèle ancestral, puisqu'il est présent de façon exclusive dans des espèces moins sélectionnées telles que le Yak ou le Buffle (Winter *et al.*, 2002). Cet allèle est marqué par un TB, un TP et une MG plus élevés et une PL et une MP plus faible que ceux observés pour l'allèle A232 (Grisart *et al.*, 2002; Winter *et al.*, 2002). L'hypothèse de la mutation K232A comme mutation causale du QTL « BTA14 » a été confirmée pour le caractère TB sur les races Holstein, Normande et Montbéliarde (Boichard *et al.*, 2003), puis Holstein allemande (Kühn *et al.*, 2004), Holstein Friesian néerlandaise et néo-zélandaise (Grisart *et al.*, 2004; Spelman *et al.*, 2002), Noires et Blanches polonaise (Pareek *et al.*, 2005) bien que l'exclusivité des deux allèles associés à cette mutation ait été remise en question (Bennewitz *et al.*, 2004). En effet, un découpage plus précis du locus *DGAT1* (Winter *et al.*, 2004), a permis d'identifier 23 gènes voisins de *DGAT1* contenant 55 positions polymorphes. Il est de plus supposé, bien que non démontré, que le nombre de répétition d'un VNTR localisé dans la région promotrice de *DGAT1* puisse être aussi impliqué dans le QTL (Gautier *et al.*, 2007; Kühn *et al.*, 2004). Bennewitz et coll. (2004), avancent l'hypothèse d'un troisième allèle explicatif pour ce QTL. Cependant, à ce jour, l'effet le plus marqué connu sur le TB reste encore la mutation K232A.

La fréquence de l'allèle K est de 0,04 ; 0,13 et 0,37 dans les populations de race Montbéliarde, Normande et Holstein française (Gautier *et al.*, 2007) alors qu'elle est de 0,6 ; 0,88 et 0,22 dans les populations Holstein-Friesian (Nouvelle-Zélande), Jersey et Ayrshire (Spelman *et al.*, 2002).

III.2.2. Du gène DGAT1 à l'enzyme fonctionnelle

La séquence de l'unité de transcription du gène *DGAT1* bovin comporte 8 101 nucléotides. Elle contient 17 exons et 16 introns de taille variable et spécifie un ARN messager (ARNm) de 1 786 nucléotides (hors queue poly(A) et avec une séquence non codante 5' partiellement déterminée). L'ARNm est traduit en une protéine de 490 acides aminés (cf. Figure III.1). L'existence de variants d'épissage a été rapportée (Grisart *et al.*, 2004). Cet épissage différentiel qui conduit à l'élimination par « intronification » d'un segment de 66 nucléotides (représentant la quasi-totalité de l'exon 8, dans sa partie 3') aboutit à une isoforme protéique de 467 acides aminés, présentant une délétion interne de 22 résidus d'acides aminés. Cette protéine semble dépourvue d'activité (Grisart *et al.*, 2004).

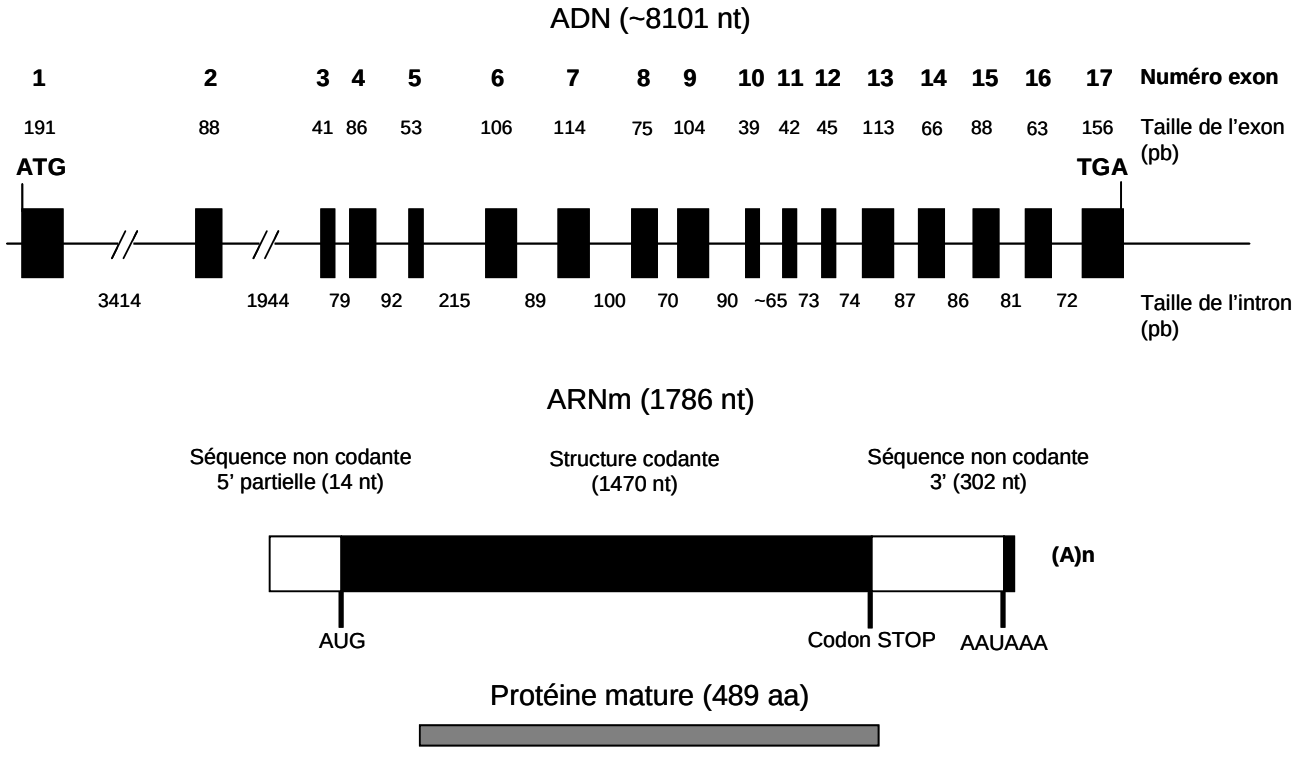


Figure III-1 : Du gène DGAT1 bovin à la protéine mature.
A partir des numéros d'accession GeneBank pour le gène, AJ318490 ; pour l'ARNm, NM_174693 et pour la protéine, NP_777118. aa : acide aminé ; ADN : acide désoxyribonucléique ; ARN : acide ribonucléique ; nt : nucléotide ; pb paire de base.

La protéine DGAT1 est une enzyme ancrée dans la membrane du réticulum endoplasmique (Wilgram & Kennedy, 1963) dont la fonction est de catalyser l'étape ultime et limitante (Mayorek *et al.*, 1989) de la synthèse des triglycérides (Figure I.22). C'est la seule enzyme qui intervient de façon exclusive dans la synthèse des triglycérides (Ozasa *et al.*, 1989). Elle catalyse l'estérification d'un acyl-CoA en position *sn-3* d'une molécule de

diacylglycérol. Son site actif se trouve sur la face cytosolique du réticulum endoplasmique (Bell *et al.*, 1981; Coleman & Bell, 1978).

Chez la souris, l'enzyme fonctionnelle a un poids moléculaire d'environ 47 kDa et peut contenir jusqu'à 9 domaines transmembranaires (Cases *et al.*, 1998). Dans le foie de rat, la taille de la structure membranaire fonctionnelle est de 72 kDa (Ozasa *et al.*, 1989) ou de 60 kDa lorsqu'elle est isolée (Andersson *et al.*, 1994). Ces masses laissent supposer que la structure fonctionnelle est composée de plusieurs chaînes polypeptidiques (Andersson *et al.*, 1994) comme démontré pour l'enzyme DGAT1 humaine pour laquelle la structure fonctionnelle serait un tétramère (Cheng *et al.*, 2001).

La spécificité de l'enzyme DGAT1 pour les substrats acyl-CoA et diacylglycérol varie selon les tissus. Ainsi, dans la muqueuse intestinale, l'enzyme DGAT1 catalyse préférentiellement des réactions au cours desquelles les diacylglycérols sont estérifiés en position sn-2 par des acides linoléiques (C18:2) ou arachidoniques (C20:4, Lehner & Kuksis, 1996). L'enzyme DGAT1 adipocytaire admet préférentiellement comme substrat des diacylglycérols contenant des acides oléiques en position sn-1 et sn-2 plutôt que des acides gras à chaîne courte ou saturée et n'admet pas d'acyl-CoA à chaîne courte comme substrat (Coleman & Bell, 1976). L'enzyme DGAT1 de l'hépatocyte de rat a une spécificité de substrat pour l'octanoate (C8:0), le décanoate (C10:0) et les AG à longue chaîne (Mayorek & Bar-Tana, 1983), son activité augmente avec les acyl-CoA C18:0, C16:0, C18:1 et C18:2 et diminue avec les acyl-CoA C18:3 et C20:4 jusqu'à être inhibée par l'acide eicosapentanoique (C20:5 *n-3*) (Rustan *et al.*, 1988).

La spécificité de substrat de l'enzyme DGAT1 du tissu mammaire bovin est caractérisée par les AG estérifiés en position *sn-3* des triglycérides du lait. Cette position est occupée en quantité décroissante par le C4:0, le C6:0, le C18:1, le C16:0 et le C18:0 (Breckenridge & Kuksis, 1968). Une étude relativement ancienne de l'activité de DGAT1, mesurée sur un tissu mammaire de vache de race Jersey en lactation montre que cette enzyme catalyse l'estérification d'AG à chaîne courte sur des 1,2-diacylglycérols à chaîne longue et qu'elle incorpore préférentiellement les acyl-CoA moyens, dans l'ordre C8:0, C16:0, C4:0 et C18:0 (Marshall & Knudsen, 1976).

Dans l'adipocyte, l'expression du gène *DGAT1* est stimulée par le glucose (Meegalla *et al.*, 2002). De plus, dans les hépatocytes de rat l'activité de l'enzyme DGAT1 est dépendante de phosphorylation/déphosphorylation (Haagsman *et al.*, 1982). Enfin, dans les tissus adipeux et le muscle bovin, son activité est optimale pour un pH proche de la neutralité (Lozeman *et al.*, 2001).

III.2.3. Effet d'une invalidation ou d'une mutation sur le gène DGAT1

L'invalidation du gène *DGAT1* chez les souris n'a pas d'incidence sur la viabilité des souris ni sur la synthèse des triglycérides. Les triglycérides sont alors synthétisés par une autre diacylglycérol acyltransférase, DGAT2 (Cases *et al.*, 2001) qui présente la même activité que l'enzyme DGAT1.

Comparativement aux souris sauvages, les souris « invalidées » pour le gène *DGAT1* (ou *Dgat1*$^{-}$/-) sont plus maigres et résistantes au « diet-induced obesity » (diabète de type 2) dû entre autres à une augmentation de la dépense énergétique et de l'activité de l'animal (Smith *et al.*, 2000). Ces souris montrent également une teneur en triglycérides moindre de leurs tissus, une sensibilité accrue à l'insuline et à la leptine (Chen *et al.*, 2002) et présentent une augmentation du métabolisme du glucose (Chen, 2006). La glande mammaire des souris femelles dgat1^{-}/- présente une incapacité à produire et sécréter du lait (Smith *et al.*, 2000), suggérant que l'enzyme DGAT1 est cruciale pour la lactation.

La mutation K232A de l'enzyme DGAT1 bovine entraîne une diminution de la vitesse maximale et donc de l'efficacité de l'enzyme entre l'allèle K (AA) et l'allèle A (GC) au locus *DGAT1* (Grisart *et al.*, 2004). Cette mutation induit également une modification de la composition du lait dans le sens d'une réduction du TB (Grisart *et al.*, 2002; Winter *et al.*, 2002) et d'une teneur moins élevée en AG saturés (surtout le C16:0) et plus élevée en C14:0, C18 insaturés et en conjugué de l'acide linoléique (C18:2) (Schennink *et al.*, 2008; Schennink *et al.*, 2007). La modification de la teneur en AG du lait due à cette mutation s'expliquerait par une altération de la spécificité de substrat de l'enzyme DGAT1 (Schennink *et al.*, 2007).

C'est dans ce contexte, qu'il a été entrepris d'étudier plus spécifiquement l'effet de la mutation K232A sur les profils d'expression des gènes dans la glande mammaire bovine en lactation. Les travaux effectués sont présentés dans la partie qui suit, sous la forme d'un projet d'article même si toutes les analyses de données qui y sont détaillées ne feront pas obligatoirement partie intégrante d'une première publication, notamment la mise en relation des données phénotypiques avec les données de transcriptomique qui feront une analyse en tant que telle.

III.3. Effet du polymorphisme K232A au locus *DGAT1* sur l'activité transcriptomique de la glande mammaire bovine : Article 2 (en préparation)

III.3.1. Introduction

La filière laitière connaît actuellement une phase critique en partie due à la baisse de consommation du lait liée à la teneur élevée en acides gras saturés de sa matière grasse (Matière Grasse Laitière, MGL), essentiellement constituée de triglycérides (97%, (Mulder & Walstra, 1974). La consommation excessive d'acides gras saturés peuvent entraîner chez l'homme des maladies cardio-vasculaires (Williams, 2000). L'une des enzymes intervenant dans la synthèse des triglycérides, et ayant une action limitante sur cette synthèse, est la Diacylglycérol Acyltransferase 1 (DGAT1). Il a été montré chez la souris que l'invalidation du gène spécifiant cette enzyme conduit à une absence totale de production laitière (Smith *et al.*, 2000). Chez les bovins laitiers, le locus *DGAT1* est localisé dans la terminaison centromérique du chromosome 14 (BTA14) où un QTL affectant la quantité de lait, de matière grasse (MG) produite et le TB a été détecté (Boichard *et al.*, 2003; Coppieters *et al.*, 1998). La différence de quantité de MG produite, mais également de taux butyreux (TB), entre individus est pour partie expliquée par la substitution de deux nucléotides successifs (AA → GC) localisée dans le huitième exon du gène DGAT1 et conduisant aux deux formes allèliques du gène les plus répandues (Grisart *et al.*, 2002). Cette double mutation induit une substitution Lysine/Alanine en position 232 de la chaîne peptidique (K232A) de la protéine qui a pour effet de modifier l'efficacité (Grisart *et al.*, 2004) et la spécificité de l'enzyme pour ses différents substrats (Schennink *et al.*, 2007). Ceci explique pour partie le changement de composition en acides gras (AG) du lait observé par (Schennink *et al.*, 2007).

Notre objectif consistait d'une part à valider l'effet de cette double mutation au locus *DGAT1* sur les performances laitières et sur la composition en AG du lait des vaches choisies à partir du dispositif mis en place au Pin-au-Haras (croisement Holstein x Normande, (Larroque *et al.*, 2003), d'autre part à étudier son effet sur la taille des globules gras, sur les caractéristiques des micelles de caséines et sur l'activité transcriptionnelle de la glande mammaire, afin de mieux cerner les autres voies métaboliques affectées par ce polymorphisme.

Pour plus de clarté, le génotype homozygote AA/AA (Lys232 ou K232) au locus *DGAT1* sera appelé génotype « fort » (en référence au critère TB) et le génotype GC/GC (Ala232 ou A232), génotype « faible ».

III.3.2. Matériels et méthodes

III.3.2.1. Animaux

Douze vaches laitières homozygotes au locus *DGAT1* ont été identifiées au sein du dispositif de détection des QTL laitiers du Pin-au-Haras sur la base de leur génotype et de leur degré de parenté. Ces animaux ont été obtenus par transfert d'embryons provenant de la 2ème génération (F2) d'un croisement Prim'Holstein x Normande (Larroque *et al.*, 2003). A chaque vache de génotype homozygote « fort » (AA/AA) au locus *DGAT1* était associé soit sa pleine sœur (n=4), soit sa demi-sœur (même père, n=2) de génotype homozygote « faible », au même stade de lactation, pour constituer *in fine* 6 couples de vaches de génotypes opposés. Les vaches homozygotes pour l'allèle *AA* (K232, génotype « fort ») étant rares dans la population expérimentale du Pin-au-Haras, nous n'avons pu constituer, en retenant tous les animaux de ce génotype que 6 couples de vaches répondant aux critères énoncés.

Ces vaches étaient issues de deux pères différents constituant ainsi deux familles (cf. Figure III.2). En 2007, les vaches retenues ont été transférées dans les Installations Expérimentales de l'URH de l'INRA de Theix, afin d'effectuer les prélèvements de lait et de tissus de la glande mammaire. Lors des prélèvements, deux couples de vaches étaient en 3ème lactation (L3), un couple en 2ème lactation (L2) et trois couples en 1ère lactation (L1).

L'originalité de ce dispositif par rapport aux études précédemment réalisées sur la mutation K232A au locus *DGAT1* réside dans le fait que les vaches d'un même couple étaient au même stade et même numéro de lactation, qu'elles étaient homozygotes et de génotype opposé au locus *DGAT1*, qu'elles se trouvaient dans un même élevage et recevaient le même régime alimentaire au moment des prélèvements. Nous nous sommes efforcés à compenser de la sorte le faible effectif par un dispositif expérimental extrêmement homogène et contrôlé.
Les animaux étaient traits à 7h00 le matin et 17h00 le soir.

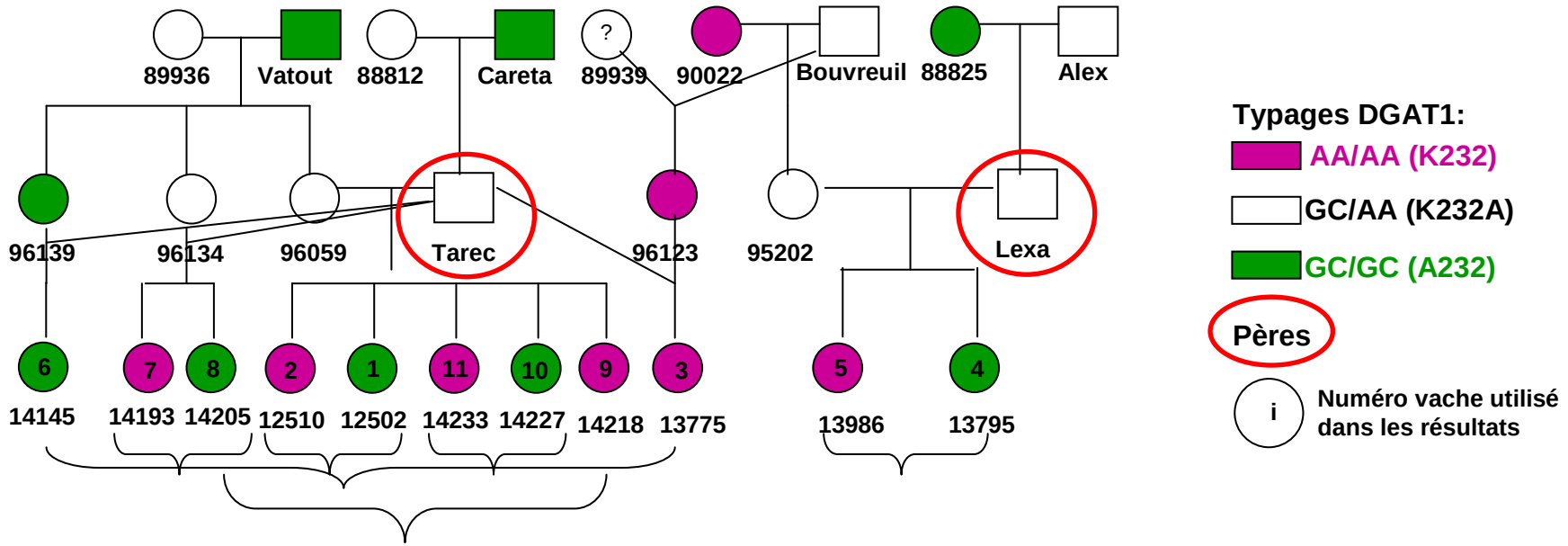


Figure III-2 : Généalogie des animaux utilisés dans l'expérimentation.
Tous les animaux ont été génotypés au locus *DGAT1*. Toutes les vaches utilisées dans l'expérimentation proviennent de deux pères différents.
Chaque accolade correspond à un couple de vaches.

III.3.2.2. Prélèvements et analyses des échantillons de lait

Des prélèvements de lait et de tissu mammaire (par biopsies) ont été effectués en moyenne à 227 ± 18 jours de lactation.

La semaine précédant les prélèvements, les performances laitières des animaux ont été mesurées tous les jours, alors que les animaux recevaient *ad libitum* un régime de type hivernal à base d'ensilage de maïs. Les variables mesurées étaient : la production laitière journalière (PL en kg/j), les taux butyreux (TB) et protéique (TP) et la teneur en lactose (TL en g/kg), la quantité de cellules dans le lait (en milliers/ml), le poids de l'animal (en kg). A partir de ces valeurs mesurées, étaient calculées les quantités de matière grasse (MG = TB x PL), protéique (MP = TP x PL) et de lactose (ML = TL x PL en g/j) secrétées.

Les analyses portant sur la matière grasse laitière (MGL) ont été effectuées sur des échantillons de lait de la traite du matin. Celles portant sur la matière protéique ont été effectuées sur des échantillons de lait de la traite du soir.

➢ Analyse sur la matière grasse laitière

L'analyse de la composition en AG des laits a été faite à l'URH de l'INRA de Theix. Celle de la taille des globules gras a été faite à l'UMR STLO de Rennes.

La composition en AG des échantillons de lait a été déterminée après extraction puis méthylation des AG des laits lyophilisés, par chromatographie en phase gazeuse sur colonne capillaire en silice fondue (100 m x 0,25 mm, Tracez-GC 2000 Series, CP'Sil 88, Chrompack), tel que décrit par Chilliard *et al.*, (2006). La teneur relative en chaque type d'AG a été déterminée à partir des chromatogrammes en calculant le rapport de l'aire du pic correspondant à l'AG à l'aire totale des pics des chromatogrammes.

Le diamètre des globules gras de chaque échantillon de lait a été mesuré à l'aide d'un Malvern Mastersizer 2000 (Malvern Instruments, Grande-Bretagne), en utilisant pour indice de réfraction de la matière grasse 1,458 à 633 nm (laser He/Ne) et 1,46 à 466 nm (diode électroluminescente) et pour indice de réfraction de l'eau 1,333. La taille des globules gras a été mesurée après dilution des échantillons de lait (1:1 vol.) avec 1 mL d'EDTA 35 mM, tamponné à pH 7,0, dans le but de dissocier les micelles de caséines (Michalski *et al.*, 2001). A partir de la distribution de la taille des globules gras (en % v/v) le logiciel du Malvern a

calculé le diamètre moyen $d_{4,3} = \Sigma(N_i \times d_i^4) / \Sigma(N_i \times d_i^3)$ (avec N_i le nombre de globules gras dans le groupe de diamètre d_i), ce calcul donnant plus de poids aux gros globules gras.

➢ Analyse sur la matière protéique

L'analyse de la composition en protéines majeures des laits a été effectuée à l'UGPL de Jouy-en-Josas. L'analyse de la taille des micelles de caséines a été faite à l'UMR STLO de Rennes.

La composition en protéines majeures des échantillons de lait a été déterminée par chromatographie en phase liquide (« Reverse Phase-High Performance Liquid Chromatography», RP-HPLC). Les protéines sont séparées selon leur hydrophobicité sur une colonne de phase apolaire d'un ProteomeLab PFD2 (Beckman Couter). Après la traite, les échantillons de lait ont été écrémés par centrifugation (5 000g, 20°C, 20 min), puis réfrigérés à 4°C pour faciliter le retrait de la couche de graisse. Ils ont ensuite été aliquotés en fractions de 1 ml conservées à -20°C. Avant injection, les échantillons étaient décongelés à température ambiante puis dilués dans 24 volumes d'une solution de clarification (0,1M bis-Tris buffer, pH à 8,0, urée 8M, citrate de sodium 1,3% et dithiothréitol 0,3%), dont le rôle est de déstabiliser les structures micellaires et de dénaturer les protéines. Les échantillons de lait clarifiés ont ensuite été filtrés sur seringue filtrante Acrodisc de 13 mm, Pall Corporation, USA (0,20 µm). Un volume de 50 µl était injecté dans le système HPLC constitué d'une colonne de 250 x 4,6 mm Jupiter C5 (Phenomenex, France) maintenue à 47°C. Les analyses en RP-HPLC ont suivi la procédure décrite par Miranda *et al.*, (2004). Les protéines sont éluées par un gradient binaire de solvant A (eau milliQ/acide trifluoroacétique (TFA) ; 1000/1 ; v/v) et de solvant B (acétonitrile/TFA ; 1000/0,8 ; v/v), préalablement filtrés à 0,45 µm (Millipore corporation, Bedford, MA). L'élution des protéines était obtenue progressivement en quatre étapes (gradients linéaires) : de 0% à 33% de solvant B pendant 10 min ; de 33% à 40% de solvant B pendant 13,5 min ; de 40% à 42% de solvant B pendant 10 min et de 42% à 52% de solvant B pendant 10 min. La colonne a ensuite été lavée avec 100% de solvant B pendant 5 min avant d'être rééquilibrée aux conditions initiales pendant 10 min. Le débit était de 1 ml/min et l'absorbance à 214 nm était enregistrée.

Les protéines et leurs variants génétiques ont été identifiés en comparant les profils chromatographiques des échantillons avec ceux obtenus sur des échantillons de lait de génotype connu au locus des protéines majeures (α-lactalbumine, β-lactoglobuline et caséines κ, β, α_{s1} et α_{s2}). Les chromatogrammes ont été intégrés sur la fenêtre 10-60 min et les pics d'intérêt ont été quantifiés (en %) relativement à l'aire totale du chromatogramme.

Pour l'analyse de la taille des micelles de caséines, les échantillons de lait ont été préalablement centrifugés 20 min à 20°C et 5 000 g et la crème éliminée. Les laits écrémés ont ensuite été filtrés sur des membranes de 0,8 µm afin d'éliminer les globules gras résiduels. La taille moyenne des micelles de caséines a été évaluée par mesure de la diffusion de la lumière à l'aide d'un Zetasizer 3000HS (Malvern Instruments, Worcestershire, UK) d'un angle de mesure de 90° et d'une longueur d'onde de 633 nm, dans le milieu dispersant de Dalgleish (20 mM d'imidazole, 50 mM de NaCl et 5 mM $CaCl_2$, pH=7) dont l'indice de réfraction et la viscosité sont respectivement de 1,333 et 0,89 cP (centipoise) à la température de 25°C.

III.3.2.3. Prélèvements et analyse des tissus mammaires

➢ Biopsie de tissus mammaires et extraction des ARN

Les biopsies de tissus mammaires ont été effectuées selon la méthode décrite par (Farr *et al.*, 1996) dans la partie supérieure de la mamelle droite non traite depuis la veille et préalablement anesthésiée. Les biopsies ont été rincées dans de l'eau physiologique (9g de NaCL dans 1 L d'eau distillée) avant d'être congelées dans de l'azote liquide, puis conservées à -80°C jusqu'à leur analyse. Les biopsies congelées ont été broyées dans un Cryobroyeur 6750 (Fisher Scientific Bioblock), sous azote liquide. Les ARN totaux ont été extraits des poudres résultantes par le réactif TRIzol en suivant les instructions du fabricant (Invitrogen, France). La phase aqueuse récupérée a ensuite été traitée à la DNase sur des colonnes Cleanup (Qiagen). Les ARN totaux ainsi extraits ont été conservés à -80°C. La quantité d'ARN a été évaluée en spectrophotométrie par détermination du ratio 260/280 nm sur le Nanodrop ND-1000 (Nanodrop Technology, Nixor Biotech, France). La qualité de chaque échantillon a été évaluée avec le Bioanalyseur 2100 (Agilent Technologies, France), en utilisant la puce Nano 6000. Les échantillons dont le RIN (« RNA Integrity Number ») était inférieur à 7,6 ont été éliminés de l'expérience.

➢ Profilage d'expression génique sur microarrays

♦ **Marquage des cibles complexes**

La préparation des cibles complexes marquées à partir des ARN totaux extraits des biopsies mammaires a été réalisée avec le kit « Low RNA Input Linear Amplification »

(Agilent Technologies, France) en suivant les instructions fournies par le fabricant. Brièvement, les ARN totaux ont été rétro-transcrits en ADNc par l'enzyme MMLV-RT. Les ADNc ont été ensuite transcrits en ARNc. Cette étape permet une amplification linéaire du nombre de molécules d'ARNc synthétisés et l'incorporation directe des cyanines liées à des CTP dans les chaînes nucléotidiques. Les ARNc marqués ont ensuite été séparés des cyanines non incorporées sur colonnes RNeasy (Qiagen, France). La fréquence d'incorporation (FOI) des cyanines (pmol/µg) et la concentration d'ARNc (pmol/µl) ont été estimées en mesurant au Nanodrop les absorbances à 260 nm, 550 nm et 650 nm. Les ratios suivants ont été calculés : FOI (Cy3) = (Abs550 x 10^3) / (6 x Abs260) ; FOI (Cy5) = (Abs650 x 100) / Abs260. Lorsque les FOI étaient inférieures à 8, les échantillons étaient écartés.

L'hybridation des cibles complexes a été réalisée selon un plan expérimental comparant deux conditions, en considérant les couples de vaches décrits précédemment (cf. Figure III.2) comme des répétitions biologiques et en effectuant des dye-swap pour les répliquas techniques.

Pour cette expérience, les cibles complexes marquées ont été hybridées, selon le même plan expérimental, sur deux types de lames à oligonucléotides (oligoarrays) bovins : les lames 4 x 44K commercialisées par Agilent (Agilent Technologies, France) et les lames 22K spottées au CRB-GADIE de Jouy-en-Josas (INRA, France).

- **Hybridation des cibles complexes sur les lames bovines 4 x 44k Agilent**

Les cibles complexes ont été hybridées sur les lames bovines 4 x 44K commercialisées par Agilent en suivant le protocole « Two Color Microarray-Based Gene Expression Array » et en utilisant le kit Agilent « Gene Expression Hybridization ». Chaque cible marquée (825 ng) était déposée sur les réseaux après fragmentation ([1x] Fragmentation Buffer 30 min à 60°C). Les lames étaient hybridées 17h à 65°C et 10 rpm dans un four à hybridation (Agilent Technologies, France). A l'issue de l'hybridation et du lavage, les cyanines fixées sur les lames ont été protégées de l'ozone par immersion des lames 30s dans une solution de stabilisation et de séchage (Agilent Technologies, France). Les lames ont été scannées sur un scanner Agilent (Agilent Technologies, France), à une résolution de 5 µm et en effectuant les acquisitions à 10 et 100 % de la puissance des lasers (option XDR, « eXtended Dynamic Range »).

♦ **Hybridation des cibles complexes sur les lames bovines 22K du CRB-GADIE**

Avant hybridation, les lames 22K bovines du CRB-GADIE (INRA, France) ont été pré-hybridées dans une solution de sérum albumine bovine (BSA, 1 g pour 100 ml) 30 min à 50°C, puis rincées dans de l'eau ultra pure avant d'être séchées. Chaque cible complexe marquée (entre 2,7 et 6 µg) était déposée sur les lames après fragmentation ([1x] Fragmentation Buffer 30 min à 60°C), en veillant à ce que la même quantité d'ARNc soit déposée sur les deux lames d'un dye-swap. Les lames ont été hybridées 20h à 60°C et 10 rpm dans un four à hybridation (Agilent Technologies, France). Après l'étape d'hybridation, les lames ont été lavées à température ambiante deux fois 5 minutes sous agitation orbitale dans une solution (2X SSC, 0.1% SDS), avant d'être successivement lavées deux fois 5 minutes dans une solution (0,5X SSC, 0,1% SDS), puis 5 minutes dans une solution (0.1X SSC, 0.1% SDS), 3 x 5 minutes dans une solution (0,1X SSC) et 1 minute dans de l'eau ultra pure, puis séchées à l'air comprimé. Les lames étaient enfin scannées sur un scanner Agilent (Agilent Technologies) avec une résolution de 10 µm et en sélectionnant l'option XDR.

L'intensité du signal fluorescent pour chaque spot de chaque image a été quantifiée au moyen du logiciel « Feature Extraction » v9.1.3.1 (Agilent Technologies, France).

➢ Quantification en PCRq

La validation des différentiels d'expression des gènes *DGAT1*, *CSN3* (caséine κ) et *CSN1S1* (caséine α_{s1}) a été effectuée par PCR quantitative en temps réel. Pour cela, 500 ng d'ARN total ont été rétrotranscrits en ADN complémentaire en présence de 0,5 µg de primer Oligo(dT) 12-18 et de SuperScript™ III (Invitrogen, France). La PCR quantitative en temps réel a été réalisée, en présence de SyBr Green, dans les conditions décrites par (Bevilacqua *et al.*, 2006). Au total, 3 gènes ont été validés par PCRq avec la protéine ribosomale S24 et la cyclophiline comme contrôles internes. Les paires d'amorces dont les séquences sont données au Tableau III.1 ont été déterminées à l'aide du logiciel « Primer Express » v2.0 (Applied Biosystems, France). Afin d'éviter l'amplification de l'ADN génomique, les amorces ont été préférentiellement positionnées sur des jonctions Exon-Exon. Les amorces ont été synthétisées par MWG Biotech (France). L'analyse des données de PCRq a été effectuée en normalisant chaque quantification par les deux gènes contrôles internes (protéine ribosomale

24S et cyclophiline) puis en calculant la moyenne des ratios pour chaque transcrit entre deux individus d'un même couple.

Tableau III-1 : Amorces des systèmes optimisés pour doser les transcrits en PCRq.
Chaque paire d'amorce amplifie un ADNc cible (la taille des amplicons variant entre 58 et 66 nucléotides). Les paires d'amorces ont été déterminées avec le logiciel « Primer Express » v2.0 (Applied Biosystems) à l'exception du système ciblant le transcrit spécifiant la protéine ribosomale 24S dont le « design » a été établi manuellement.

Gènes	Amorces	Séquence 5'_3'	Taille de l'amplicon
Caséine α_{s1}	Forward	TCC ACT AGG CAC ACA ATA CAC TGA	61 nt
(*CSN1S1*)	Reverse	GCC AAT GGG ATT AGG GAT GTC	
Caséine κ	Forward	AGG TGC AAT GAT GAA GAG TTT TTT C	66 nt
(*CSN3*)	Reverse	CCC AAA AAT GGC AGG GTT AA	
DGAT1	Forward	GGC GGT CCC CAA CCA	58 nt
	Reverse	GCA GGA GTG GAA GAG CCA GTA	
Cyclophiline	Forward	TGA CTT CAC ACG CCA TAA TGG T	
	Reverse	CAT CAT CAA ATT TCT CGC CAT AGA	
Protéine ribosomale 24S	Forward	TTT GCC AGC ACC AAC GTT G	66 nt
	Reverse	AAG GAA CGC AAG AAC AGA ATG AA	

III.3.2.4. Analyse des données

➢ Analyse statistique des données phénotypiques

Les données phénotypiques (zootechnie et mesures sur les structures supramoléculaires du lait) ont été traitées par analyse de variance en utilisant la fonction *lme* du package « nlme » développé sous R par J. Pinheiro et D. Bates. Le modèle mixte appliqué aux variables phénotypiques était :

$$\mathbf{Y = \gamma_g + \lambda_l + (\gamma*\lambda)_{gl} + P_p + E_{glp}}$$

avec Y la variable phénotypique analysée, γ_g l'effet fixe du génotype g (n=2) ; λ_l l'effet fixe de la lactation l (n=3) ; $(\gamma*\lambda)_{gl}$ l'effet de l'interaction du génotype g avec la lactation l (n=6) ; P_p l'effet aléatoire du père p (n=2) et E_{glp} l'erreur résiduelle du modèle.

Lorsque l'interaction entre le génotype et la lactation était significative, les données étaient de nouveau analysées en utilisant la même procédure mais en retirant du modèle l'interaction, afin de valider l'effet du génotype ou de la lactation sur le caractère analysé.

Dans le modèle d'analyse, nous avons considéré l'effet père comme aléatoire, puisqu'il était difficile d'intégrer dans le modèle l'association privilégiée entre deux vaches (pleine ou demi-sœurs).

➢ Analyse en Composantes Principales (ACP) sur les données phénotypiques

Afin de visualiser les corrélations entre variables phénotypiques mesurées sur les individus, une analyse en composantes principales (ACP) a été réalisée sur les données de performances laitières, de taille des globules gras et des micelles de caséines, des teneurs en AG et de la teneur en caséine α_{s1}. Cette ACP a été effectuée en utilisant la fonction « *PCA* » du package « FactoMineR » de R (Lê *et al.*, 2008). Les variables qualitatives des individus, retenues pour cette analyse, étaient le père (n=2), le génotype au locus *DGAT1* (n=2) et le stade de lactation (n=3).

➢ Analyses des données issues des puces à ADN

♦ Réannotation des lames

Avant d'effectuer une analyse sur les données issues des expériences de profilage expressionnel, les sondes (séquences oligonucléotidiques) présentes sur les lames du CRB et d'Agilent ont été réannotées. Pour ce faire, chaque séquence a été associée à un numéro de contig Sigenae (http://www.sigenae.org/), permettant de récupérer via « Biomart » une annotation GeneBank bovine ou Unigene humaine.

♦ Normalisation, vérification de l'ensemble des données et concaténation des données des deux types de lame

Les données issues des deux types de lame (CRB et Agilent) ont été normalisées séparément. Cette normalisation a été effectuée en utilisant la fonction spécifiée « *normalisation* » du package « anapuce » sous le logiciel R qui procède par une normalisation Lowess suivie, uniquement pour les lames du CRB, d'une correction des signaux par bloc. La cohérence du fichier des signaux normalisés a été vérifiée en utilisant la fonction *verifdonnee* développée sous R. Cette fonction calcule (i) les coefficients de corrélation entre les lames, pour vérifier que les swaps présentent des signaux opposés, (ii) les coefficients de corrélation entre les intensités des signaux normalisés de toutes les lames, pour vérifier que l'effet lame n'est pas plus fort que l'effet échantillon et que les répétitions biologiques d'une même condition sont bien corrélées. Finalement, cette fonction trace des box-plot permettant de visualiser rapidement si les lames sont comparables entre elles.

Après vérification des ensembles de données normalisés issus des deux types de lame, nous avons décidé de concaténer les deux jeux pour former une « super lame » contenant à la

fois les sondes de la lame CRB et celles de la lame Agilent. Le nouveau jeu de données a été entièrement revérifié avec la fonction *verifdonnee*.

Cette concaténation a été réalisée dans l'objectif de simplifier l'analyse (un seul fichier à analyser) et d'effectuer la correction de l'erreur du test multiple (« False Discovery Rate », FDR) sur toutes les sondes disponibles et exploitées dans l'expérience.

♦ Analyse différentielle

Cette « super lame » a été soumise à une analyse différentielle en utilisant la fonction *DiffAnalysis* du package « anapuce » sous le logiciel R. Les options choisies étaient de classer les gènes par groupe de variance homogène (Delmar *et al.*, 2005), de corriger les probabilités critiques par la méthode de Benjamini et Hochberg (Benjamini & Hochberg, 1995) et de calculer la moyenne des informations des spots correspondant à la même sonde.

De plus, le niveau d'expression des gènes a été évalué, pour chaque génotype, en calculant l'intensité du signal dans chaque canal (Cy3 et Cy5). Un gène a été considéré comme (i) fortement exprimé lorsque son intensité était supérieure à 10% de l'intensité la plus élevée observée dans l'ensemble des données, (ii) faiblement exprimé lorsque son intensité était inférieure à dix fois l'intensité la plus faible, (iii) moyennement exprimé lorsque son intensité était comprise entre les deux critères précédents. Nous avons choisis ces critères de classification après observation du signal de plusieurs spots sur les lames.

♦ Interprétation biologique

Pour déterminer les principales fonctions métaboliques affectées par le polymorphisme au locus *DGAT1*, la liste des gènes différemment exprimés a été importée dans le logiciel « Ingenuity Pathway Analysis » (IPA) v6.0 (Ingenuity® Systems, www.ingenuity.com) et l'interprétation biologique s'est faite à partir des résultats issus des analyses effectuées avec ce logiciel.

➢ Mise en relation des données phénotypiques avec les données de transcriptomique

Afin de mettre en relation les données phénotypiques et les données de transcriptomique, une analyse en composantes principales (ACP) a été effectuée sur l'intensité moyennée par animal du signal des gènes déclarés différemment exprimés entre les deux génotypes au locus *DGAT1*, en utilisant la fonction « *PCA* » du package « FactoMineR » de R (Lê *et al.*, 2008). Les variables qualitatives des individus retenues pour cette analyse étaient le père (n=2), le génotype au locus *DGAT1* (n=2) et le stade de lactation (n=3).

Ensuite, les coordonnées des individus sur l'axe qui discriminait le génotype au locus *DGAT1* ont servi de variable qualitative pour une analyse de variance réalisée sur les données phénotypiques avec la fonction « *aov* » du package « stats » de R. Nous avons pu ainsi déterminer si les coordonnées de l'axe avaient un effet significatif sur les variables phénotypiques analysées.

III.3.3. Résultats

III.3.3.1. Données phénotypiques

- **Effet du polymorphisme au locus *DGAT1* sur les performances laitières et les caractéristiques physico-chimiques des laits**

Pour déterminer si le génotype au locus *DGAT1* exerce un effet significatif sur les performances laitières des animaux et sur la taille des structures supramoléculaires du lait (globules gras et micelles de caséines), une ANOVA a été appliquée sur les données en paramétrant l'effet père en effet aléatoire.

Dans notre étude, le génotype au locus *DGAT1* n'a pas d'effet significatif sur la production laitière, le taux protéique, la teneur en lactose ni sur les quantités de matières sécrétées (matières grasse et protéique et lactose). La quantité de lactose sécrétée a toutefois tendance à être plus élevée avec le génotype faible au locus *DGAT1* (949 g/j contre 769 g/j, cf. Tableau III.2). La taille des micelles de caséines n'est pas affectée par le polymorphisme au locus *DGAT1*. En revanche, les globules gras sont en moyenne significativement plus petits dans le lait des individus de génotype « faible » (4,3 µm contre 4,9 µm, avec $P<0,01$ après élimination de l'effet de l'interaction entre le génotype et la lactation).

Les individus de génotype « faible » se caractérisent donc par un TB plus faible (41,6 g/kg contre 51,6 g/kg), des globules gras de taille significativement plus réduite et une tendance (non significative, mais marquée) à une augmentation de la quantité de lactose sécrétée.

Tableau III-2 : Effets du génotype au locus *DGAT1* sur les performances zootechniques, la taille des globules gras et des micelles de caséines
(ML : quantité de lactose secrétée, MG : matière grasse, MP : matière protéique, NS : effet non significatif à 5%, Pc : probabilité critique, PL : production laitière, TB : taux butyreux, TGG : taille globules gras, TL : teneur en lactose, TMC : taille micelles de caséines, TP : taux protéique, * effet de l'interaction génotype x lactation, σ : écart-type)

	Génotype "Fort" AA/AA, K232 (σ)	Génotype "Faible" GC/GC, A232 (σ)	Pc
PL (kg/j)	16,0 (2,7)	19,3 (6,3)	NS
TB (g/kg)	**51,6 (2,8)**	**41,6 (2,4)**	**P<0,001**
TP (g/kg)	34,5 (2,9)	32,6 (1,3)	NS
TL (g/kg)	47,8 (3,9)	49,6 (2,2)	NS
Cellules (milliers/ml)	206,3 (291,3)	255,7 (354,6)	NS*
Poids (kg)	681,3 (46,3)	670,2 (58,3)	NS
MG (g/j)	826,2 (159,8)	792,5 (221,8)	NS
MP (g/j)	553,9 (119,4)	625,8 (196,9)	NS
ML (g/j)	768,5 (167,3)	948,7 (275,9)	NS
TGG (µm)	**4,9 (0,5)**	**4,3 (0,2)**	**P<0,01***
TMC (nm)	167,1 (12,0)	164,9 (13,2)	NS

- **Effet du polymorphisme au locus *DGAT1* sur la composition en acides gras des laits**

La composition en AG des laits prélevés deux jours avant la biopsie des tissus mammaires a été déterminée par chromatographie en phase gazeuse. L'effet du génotype au locus *DGAT1* sur la composition en AG des laits a été déterminé en appliquant une ANOVA, l'effet « père » étant aléatoire. Les laits des individus de génotype « faible » contenaient significativement moins de C14:1 et C16:1 *cis*9 et plus de C14:0, C18:1 *cis*11, C18:1 *cis*12, C18:2 *n-6*, C18:3 *n-3* que les laits des individus de génotype « fort » (Tableau III.3). Les index de désaturation des acides caprique (C10:0), myristique (C14:0) et palmitique (C16:0) étaient significativement plus faibles avec le génotype « faible » au locus *DGAT1*. La somme des acides caprique, laurique (C12 :0) et myristique est significativement plus élevée pour le génotype « faible » au locus *DGAT1*.

Ainsi, les individus de génotype « faible » se caractérisent par un lait plus riche en AG saturés à chaîne moyenne (de C10:0 à C14:0) et en AG longs et insaturés.

Tableau III-3 : Effet du polymorphisme au locus DGAT1 sur la composition en acides gras des laits

Acide gras (%)	Génotype "Fort" AA/AA, K232 (σ)	Génotype "Faible" GC/GC, A232 (σ)	Pc
C14:0	12,74 (1,07)	13,61 (0,40)	P<0,1
C14:1	1,19 (0,07)	0,992 (0,08)	P<0,01
C16:1 *cis9*	1,54 (0,14)	1,15 (0,13)	P<0,01
C18:1 *cis11*	0,47 (0,10)	0,57 (0,04)	P<0,1
C18:1 *cis12*	0,17 (0,04)	0,22 (0,03)	P<0,1
C18:2 n-6	1,13 (0,14)	1,44 (0,21)	P<0,05
C18:3 n-3	0,27 (0,04)	0,35 (0,03)	P<0,01
s10et12et14*	21,41 (2,31)	22,88 (1,23)	P<0,1
index10**	9,06 (1,06)	7,84 (0,25)	P<0,05
index14**	8,49 (0,80)	6,73 (0,49)	P<0,01
index16**	4,26 (0,49)	3,44 (0,10)	P<0,01

* s10et12et14 = C10:0 (%) + C12:0 (%) + C14:0 (%)

** Les index de désaturation sont calculés comme indiqué pour l'exemple suivant :
index14 = C14:1 *cis9*/(C14:1 *cis9*+ C14:0)

σ = écart-type, Pc = probabilité critique

♦ Détermination des variants protéiques et de la composition en protéines majeures des laits

La séparation des différentes protéines et de leurs variants dans les conditions RP-HPLC utilisées dans cette expérience, n'a pas été totalement concluante, en particulier pour les protéines majeures du lactosérum. En effet, l'α-lactalbumine n'a pu être individualisée du variant B de la β-lactoglobuline (Figure III.3). En revanche, la séparation des variants A et B de la caséine κ, des deux variants de la caséine α_{s2} et des quatre variants de la caséine β, est excellente. Il convient de noter que dans ces conditions les variants B et C de la caséine α_{s1}, ne peuvent être distingués. Le génotype au locus *CSN1S1*, spécifiant la caséine α_{s1}, n'a pas été déterminé sur les animaux de cette étude.

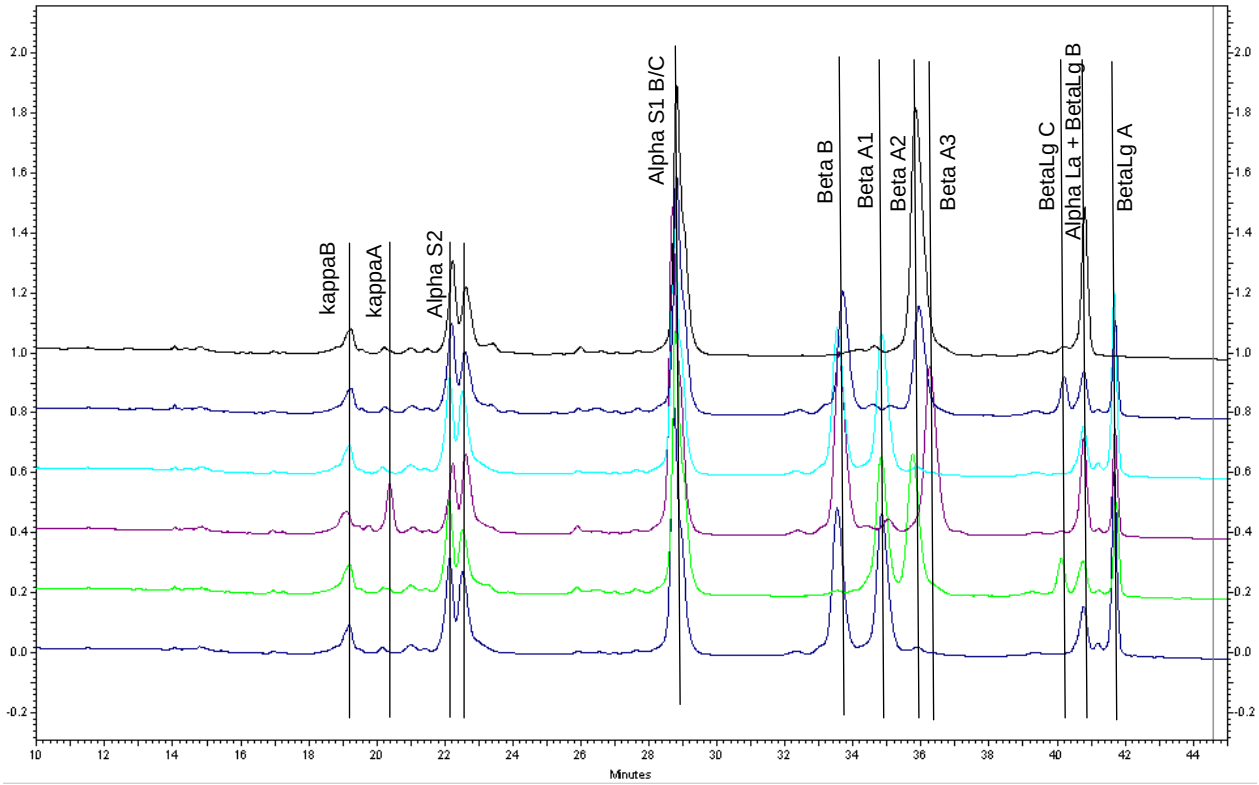


Figure III-3 : Analyse de la fraction « protéines majeures » des laits en RP-HPLC.
kappaB = caséine κ variant B, kappaA = caséine κ variant A, Alpha S2 = caséine α_{s2}, Alpha S1 B/C = caséine α_{s1} variant B et/ou C, Beta B = caséine β variant B, Beta A1 = caséine β variant A1, Beta A2 = caséine β variant A2, Beta A3 = caséine β variant A3, BetaLg C = β -lactoglobuline variant C, AlphaLa + BetaLgB = α–lactalbuline et β -lactoglobuline variant B, BetaLg A = β -lactoglobuline variant A.

Le Tableau III.4 indique, pour chaque animal, les phénotypes pour les protéines majeures.

Tableau III-4 : Phénotypage (variants protéiques) des animaux du dispositif pour la β-lactoglobuline et les caséines κ et β par analyse des laits écrémés en RP-HPLC.

Génotype au locus *DGAT1*	Numéro de l'animal	Caséine κ	Caséine β	β-lactoglobuline
Génotype « Faible » GC/GC, A232	1	AB	BA2	AB
	4	AA	A1A2	BB
	10	AB	A2A2	AB
	8	BB	A2A2	AB
	6	BB	BA1	AA
Génotype « Fort » AA/AA, K232	2	AB	BA2	AB
	5	BB	BB	AB
	11	AB	A2A2	AA
	7	BB	BA1	AB
	3	AB	A1A2	AA
	9	AB	A3A3	AA

Sur la base de ces analyses phénotypiques, on n'observe pas d'individu homozygote AA pour la caséine κ et pas davantage d'homozygote BB pour la β-lactoglobuline dans le groupe « génotype fort » au locus *DGAT1* et il n'y a pas d'homozygote A3A3 et BB pour la caséine β dans le groupe « génotype faible ». De plus, tous les individus semblent hétérozygotes pour la caséine α_{s2}, sauf un individu de génotype *DGAT1* « faible ». Or, le polymorphisme génétique d'une protéine peut avoir un impact sur la quantité de cette protéine (pour revue Martin *et al.*, 2002). Il est donc impossible, dans notre étude, de séparer l'effet du génotype au locus *DGAT1* des effets des génotypes au loci spécifiant les lactoprotéines majeures. Nous devons souligner ici les limites de notre dispositif qui comporte un effectif insuffisant d'animaux. Cela étant dit, seuls les loci *BLG* et *CSN3* ont à ce jour pu être associés, en bovin, à des variations d'expression en liaison avec les formes alléliques qu'ils présentent.

- **Analyse en composante principale sur les données phénotypiques**

Les corrélations entre les performances laitières, la taille des globules gras et les teneurs pour les acides gras qui varient selon le génotype au locus *DGAT1* ont été visualisées à l'aide d'une analyse en composante principale (ACP).

Les deux premiers axes de l'ACP expliquaient environ 67% de la variabilité. Dans ce plan, la population d'individus de génotype « fort » était plus dispersée que celle de génotype « faible ». En fixant le seuil de décision pour conserver une variable à 5% (option proba = 0,05 de la fonction « *PCA* » du package « FactoMineR »), c'est la deuxième dimension qui discrimine significativement les animaux en fonction du génotype au locus *DGAT1*. Les autres variables qualitatives (père et stade de lactation) n'étaient pas discriminées à ce seuil. Au seuil de 5%, la première dimension n'était pas caractérisée par les variables qualitatives du modèle (génotype, père et lactation).

Le côté positif de la deuxième dimension représente les individus de génotype « faible » (GC) et le côté négatif les individus de génotype « fort » (AA) (cf. Figure III.4). Sur cet axe, se répartissaient dans le sens positif (équivalent au génotype « faible ») les variables quantitatives teneurs en AG C14:0, C18:1 *trans*4 et somme du C10:0, C12:0 et C14:0, le TL et la teneur en lactose. Dans le sens négatif du deuxième axe (représenté par le génotype « fort ») se répartissaient les variables qualitatives taille du globule gras, TB, teneurs en C14:1 et C16:1 *cis*9 et les index de désaturation du C10:0, C14:0 et C16:0.

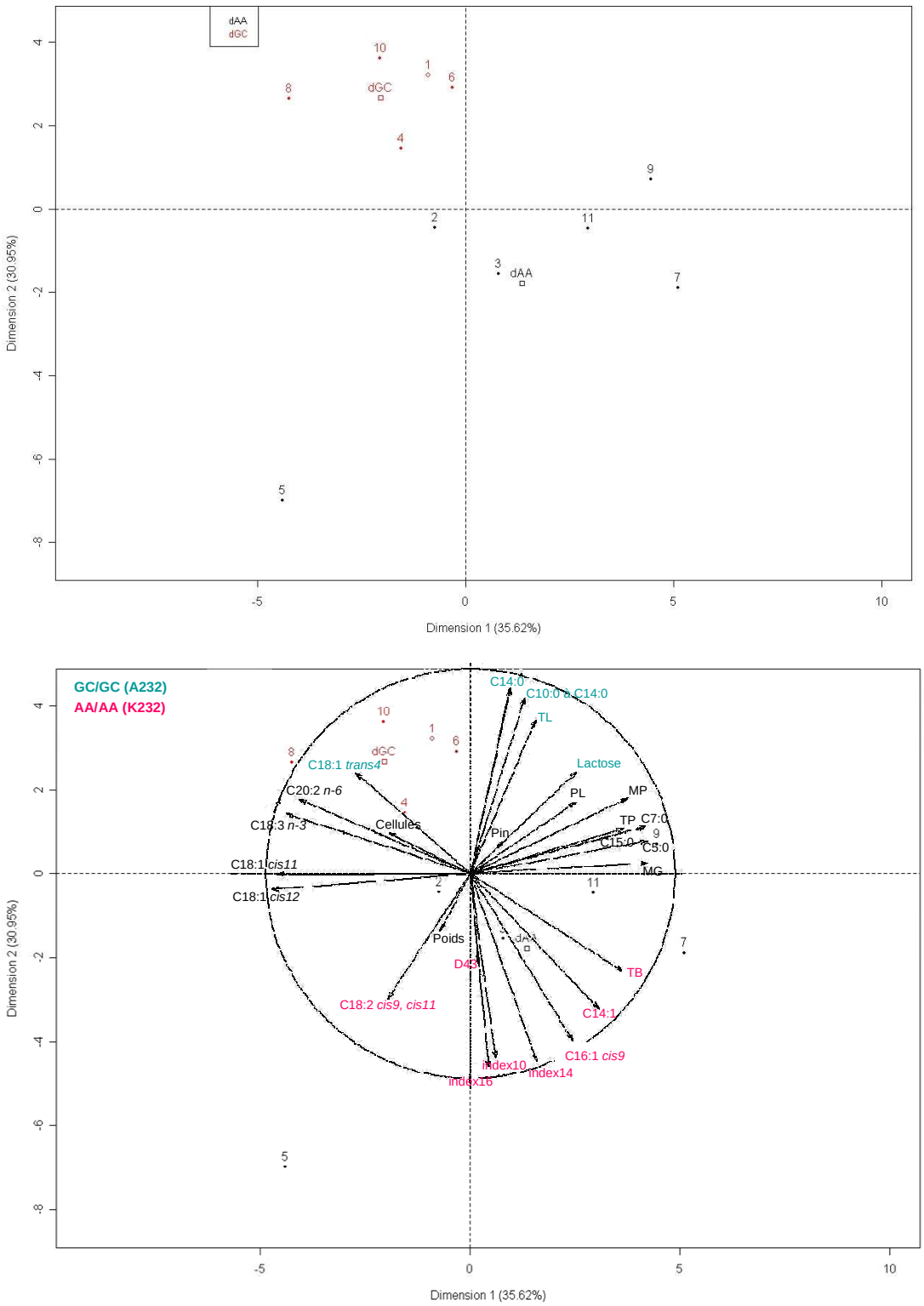


Figure III-4 : Analyse en composante principale sur les variables phénotypiques.
A : répartition des individus dans le plan contenant les deux premières dimensions (67% de la variabilité expliquée) - B : répartition des variables qualitatives dans le plan contenant les deux premières dimensions. dGC = génotype « faible » au locus *DGAT1* GC/GC, A232 ; dAA = génotype « fort » au locus *DGAT1* AA/AA, K232. En vert les variables caractérisant les individus de génotype « faible ». En rouge les variables caractérisant les individus de génotype « fort ». D43 = taille des globules gras

Les individus de génotype « faible » se caractérisent donc par une teneur en AG C14:0, C18:1 *trans*4 et somme du C10:0, C12:0 et C14:0, une teneur en lactose et un TL, plus élevés. Les individus de génotype « fort » se caractérisent par une taille des globule gras, un TB, des index de désaturation pour C10:0, C14:0 et C16:0 et une teneur en C14:1 et C16:1 *cis*9, plus élevés. Les autres variables quantitatives ne se distribuaient pas sur cet axe, avec un seuil de décision fixé à 5%.

➢ Résultats de l'analyse trancriptomique

Après normalisation, la cohérence des deux jeux de données provenant des deux types de lame ont été vérifiés en appliquant la fonction « *verifdonnee* » développée sous R, basée sur les corrélations entre lames et entre intensités du signal pour chaque individu. Après vérification, ces deux ensembles de données ont été concaténés. Le nouvel ensemble a été entièrement revérifié (cohérence) et est équivalent à une « super lame » comportant 43 061 séquences oligonucléotidiques (sondes) différentes qui correspondent à 26 934 contigs « Sigenae » différents, soit un taux de recouvrement entre les lames CRB et Agilent de 25%, dont 14 567 se rapportent à des gènes différents (soit 55% du génome bovin) ayant une annotation acceptée par le logiciel d'interprétation biologique IPA.

Pour déterminer les gènes différemment exprimés entre les deux génotypes au locus *DGAT1*, une analyse différentielle avec un modèle de variance mixte (Delmar *et al.*, 2005), basée sur la fonction « *DiffAnalysis* » du package « anapuce » de R, a été appliquée sur l'ensemble des données. Cette méthode construit des groupes de gènes. Dans chaque groupe, les gènes ont une variance homogène. La multiplicité du test a été prise en compte, en ajustant les probabilités avec la procédure de Benjamini-Hochberg (Benjamini & Hochberg, 1995) qui contrôle les faux positifs. Comme la fonction « *DiffAnalysis* » est figée sur une comparaison de deux conditions, et ne permet pas de mettre une variable en aléatoire, nous avons essayé une analyse de variance gène à gène en modélisant l'effet « père » en aléatoire. Cela n'a en rien modifié l'ordre des gènes déclarés différemment exprimés. Nous avons donc décidé de rester sur l'analyse différentielle utilisant la fonction « *DiffAnalysis* » du package « anapuce ».

Avec un seuil de significativité fixé à 10%, 1 991 sondes (soit 5% des sondes présentes dans le jeu de données) présentaient une intensité du signal différente entre les deux génotypes au locus *DGAT1*. Ces sondes correspondaient à 1 496 gènes différents ayant une annotation acceptée par IPA. 984 sondes (soient 698 gènes différents) étaient surexprimées et 1 007 sondes (soient 798 gènes différents) étaient sous exprimées, pour le génotype « faible ». Devant le nombre important de gènes à analyser, nous avons fixé une limite supplémentaire

en introduisant un différentiel d'expression de 1,5 (en plus ou en moins). En imposant cette limite, nous avons restreint à 270 « sondes » (soit 197 gènes) surexprimées et 268 « sondes » (soit 228 gènes) sous-exprimées pour le génotype « faible ». Nous avons ensuite vérifié que ce panel de gènes était associé aux mêmes fonctions métaboliques que le pool de gènes déclarés différemment exprimés au seuil 10% sans limite dans la grandeur du différentiel d'expression.

Les gènes surexprimés pour le génotype « faible » au locus *DGAT1* caractérisent les fonctions biologiques : métabolisme lipidique, biochimie des petites molécules et métabolisme des glucides (cf. Tableau III.5). Certains gènes caractérisent une seule fonction biologique, d'autres participent à toutes les fonctions biologiques.

La majorité des gènes surexprimés chez le génotype « faible » au locus *DGAT1* spécifient des protéines spécialisées dans la voie de synthèse des lipides du lait. Dans cette voie métabolique, la captation des lipides par la cellule épithéliale mammaire était représentée par les gènes spécifiant la lipoprotéine lipase (*LPL*) et la CD36. La synthèse des AG est corrélée à la surexpression de l'acétyl-CoA carboxylase (*ACACA*) et leur transport intracellulaire est associé aux « fatty acid binding proteins » 3 et 4 (*FABP3*, *FABP4*). La synthèse des diacylglycérols est révélée par la surexpression des gènes spécifiant la glycolyse kinase (*GK*), l'1-acylglycérol-3-phosphate O-acyltransférase 6 (*AGPAT6*) et la glycérol-3-phosphate acyltransférase (*GPAM*). La lécithine rétinol acyltransférase (*LRAT*) et la phosphatidyléthanolamine N-méthyltransférase (*PEMT*) spécifient des protéines qui participent à la synthèse des lipides membranaires. Les autres gènes surexprimés spécifient des protéines impliquées : (i) dans la sécrétion lipidique, c'est le cas de l' « ATP-binding cassette, sub-family G, member 2 » (*ABCG2*), de la butyrophiline (*BTN1A1*) et de la mucine 1 (*MUC1*) et (ii) dans la régulation de la synthèse et de la sécrétion des lipides telle le « peroxisome proliferator-activated receptor gamma, co-activator 1 alpha » (*PPARGC1A*).

Certains gènes spécifiant les protéines du lait étaient également surexprimés avec le génotype « faible » au locus *DGAT1*. C'était le cas pour l'α-lactalbumine (*LALBA*) qui participe aussi à la synthèse du lactose et les caséines α_{s1} et α_{s2} (*CSN1S1* et *CSN1S2*, données non montrées).

Tableau III-5 : Fonctions métaboliques déterminées par « Ingenuity Pathway Analysis » sur les gènes significativement surexprimés avec le génotype « faible » (GC/GC, A232) au locus *DGAT1*

Symbole du gène	Protéine codée par le gène	Ln2(R/V)	Ratio d'expression	Numéro de la sonde	Numéro d'accession	Intensités du signal
Gènes sur exprimés pour le génotype GC/GC, A232						**A232/K232**
Biochimie des petites molécules						
ALDH18A1	aldehyde dehydrogenase 18 family, member A1	0,59	1,51	A_73_100261	XM_872596	moyen/moyen
CTNS	cystinosis, nephropathic	0,72	1,65	A_73_120818	XM_864412	moyen/moyen
DIO2	deiodinase, iodothyronine, type II	0,75	1,68	A_73_113164	CB456458	moyen/moyen
ENPP3	ectonucleotide pyrophosphatase/phosphodiesterase 3	0,63	1,54	A_73_114715	XM_607852	moyen/moyen
ENTPD4	ectonucleoside triphosphate diphosphohydrolase 4	0,61	1,53	A_73_115993	AW430479	moyen/faible
FAH	fumarylacetoacetate hydrolase (fumarylacetoacetase)	0,77	1,70	A_73_111280	XM_585546	moyen/moyen
FOLR1	folate receptor 1 (adult)	0,98	1,97	A_73_104769	XM_589186	moyen/moyen
LPO	lactoperoxidase	0,93	1,90	A_73_120322	NM_173933	fort/fort
MTHFR	5,10-methylenetetrahydrofolate reductase (NADPH)	0,88	1,85	A_73_108814	NM_001011685	moyen/moyen
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	0,65	1,57	A_73_101570	XM_869019	faible/faible
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	0,60	1,51	OLIGO_03338	CN441227	faible/faible
PAH	phenylalanine hydroxylase	0,87	1,83	A_73_104247	XM_587738	moyen/moyen
PAPSS1	3'-phosphoadenosine 5'-phosphosulfate synthase 1	0,68	1,60	A_73_105307	NM_001034210	moyen/moyen
PAPSS1	3'-phosphoadenosine 5'-phosphosulfate synthase 1	0,66	1,58	Bt00002409	CK773516	moyen/faible
SLC16A1	solute carrier family 16, member 1 (monocarboxylic acid transporter 1)	0,86	1,82	Bt00005209	CK778726	faible/faible
SLC1A2	solute carrier family 1 (glial high affinity glutamate transporter), member 2	0,70	1,63	A_73_100861	CB448035	moyen/moyen
SLC25A4	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	0,66	1,58	A_73_119975	NM_174658	moyen/moyen
SLC34A2	solute carrier family 34 (sodium phosphate), member 2	1,62	3,07	A_73_118765	NM_174661	moyen/moyen
SLC34A2	solute carrier family 34 (sodium phosphate), member 2	0,60	1,52	A_73_105601	NM_174661	faible/faible
SLC7A5	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	0,98	1,97	A_73_108397	NM_174613	moyen/moyen
SLC7A5	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5	0,92	1,89	OLIGO_07994	DN512890	moyen/moyen
Métabolisme des glucides						
HPSE	heparanase	1,17	2,25	A_73_104774	NM_174082	moyen/moyen
Métabolisme lipidique et biochimie des petites molécules						
ABCD1	ATP-binding cassette, sub-family D (ALD), member 1	0,63	1,55	A_73_102327	XM_593152	moyen/moyen
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	1,14	2,20	Bt00007511	CK941351	moyen/moyen
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	1,10	2,14	A_73_111728	NM_001037478	fort/moyen
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	1,09	2,13	Bt00006951	CK848971	moyen/moyen
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	0,94	1,92	OLIGO_07799	AW462521	moyen/faible
ACACA	acetyl-Coenzyme A carboxylase alpha	0,70	1,63	A_73_111542	NM_174224	moyen/moyen
ADFP	adipose differentiation-related protein	0,63	1,55	A_73_115567	NM_173980	fort/fort
ADFP	adipose differentiation-related protein	0,62	1,54	OLIGO_06432	CV976041	moyen/moyen
AGPAT6	1-acylglycerol-3-phosphate O-acyltransferase 6 (lysophosphatidic acid acyltransferase, zeta)	0,84	1,79	A_73_119291	XM_866018	fort/fort
ALOX15	arachidonate 15-lipoxygenase	1,49	2,81	A_73_102134	NM_174501	fort/moyen
ANG	angiogenin, ribonuclease, RNase A family, 5	0,85	1,80	A_73_112116	NM_001040590	moyen/moyen
BTN1A1	butyrophilin, subfamily 1, member A1	0,72	1,65	OLIGO_12085	M35551	moyen/moyen
BTN1A1	butyrophilin, subfamily 1, member A1	0,70	1,63	A_73_114446	NM_174508	fort/fort
CIDEA	cell death-inducing DFFA-like effector a	0,94	1,92	A_73_100290	XM_605438	fort/fort
CIDEA	cell death-inducing DFFA-like effector a	0,93	1,90	Bt00007235	CK849502	moyen/moyen
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	1,10	2,15	A_73_107233	XM_588298	moyen/faible
CYP26A1	cytochrome P450, family 26, subfamily A, polypeptide 1	0,94	1,92	A_73_112278	XM_584484	faible/faible
GFRA2	GDNF family receptor alpha 2	1,28	2,42	A_73_113101	NM_001015595	moyen/moyen
MUC1	mucin 1, cell surface associated	0,65	1,57	A_73_115084	NM_174115	fort/fort
NCOA4	nuclear receptor coactivator 4	0,73	1,66	A_73_109490	XM_603681	moyen/moyen
PCCB	propionyl Coenzyme A carboxylase, beta polypeptide	0,60	1,51	A_73_119213	NM_001038548	moyen/moyen
RETSAT	retinol saturase (all-trans-retinol 13,14-reductase)	1,16	2,23	A_73_100979	XM_865332	moyen/faible
THRB	thyroid hormone receptor, beta (erythroblastic leukemia viral (v-erb-a) oncogene homolog 2, avian)	0,92	1,89	A_73_106879	BM031001	moyen/moyen
Biochimie des petites molécules et métabolisme des glucides						
BSG	basigin (Ok blood group)	0,59	1,50	A_73_109550	XM_585535	moyen/moyen
SLC5A1	solute carrier family 5 (sodium/glucose cotransporter), member 1	0,69	1,61	A_73_102890	NM_174606	moyen/moyen
TSTA3	tissue specific transplantation antigen P35B	0,67	1,59	A_73_105235	XM_590801	moyen/moyen
Métabolisme lipidique, biochimie des petites molécules et métabolisme des glucides						
ADIPOR2	adiponectin receptor 2	0,77	1,71	A_73_104090	BM087193	moyen/moyen
CD36	CD36 molecule (thrombospondin receptor)	1,00	2,00	A_73_108146	BF603212	moyen/moyen
CD36	CD36 molecule (thrombospondin receptor)	0,80	1,74	A_73_103328	NM_174010	faible/faible
FABP3	fatty acid binding protein 3, muscle and heart (mammary-derived growth inhibitor)	0,71	1,63	A_73_111582	NM_174313	fort/fort
FABP4	fatty acid binding protein 4, adipocyte	0,79	1,73	A_73_113342	CB537957	moyen/moyen
GK	glycerol kinase	0,63	1,55	A_73_107388	XM_582365	faible/faible
GPAM	glycerol-3-phosphate acyltransferase, mitochondrial	0,92	1,89	OLIGO_13062	NM_001012282	moyen/faible
GRK4	G protein-coupled receptor kinase 4	0,59	1,50	A_73_118972	XM_590012	moyen/moyen

LALBA	lactalbumin, alpha-	1,15	2,22	A_73_113728	NM_174378	fort/fort
LPL	lipoprotein lipase	0,90	1,87	A_73_108260	XM_616349	fort/fort
LRAT	lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase)	1,02	2,03	A_73_107724	NM_177503	moyen/moyen
LRAT	lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase)	0,94	1,93	A_73_117341	NM_177503	moyen/faible
PEMT	phosphatidylethanolamine N-methyltransferase	0,83	1,78	A_73_102178	NM_182989	faible/faible
PPARGC1A	peroxisome proliferator-activated receptor gamma, coactivator 1 alpha	1,70	3,24	A_73_108828	NM_177945	moyen/moyen
PRKAR2B	protein kinase, cAMP-dependent, regulatory, type II, beta	0,65	1,57	A_73_112893	AW357054	moyen/moyen
PTHLH	parathyroid hormone-like hormone	1,48	2,80	A_73_100098	NM_174753	moyen/moyen

Les gènes sous-exprimés pour le génotype « faible » caractérisent les fonctions métaboliques : signalisation et interaction cellulaire, mort cellulaire, mouvement cellulaire et prolifération et croissance cellulaire (cf. Tableau III.6). Ces gènes spécifient des protéines qui peuvent être classées en plusieurs catégories :

1 - les protéines du cytosquelette, représentées par les kératines 4, 6A, 8 et 18 (*KRT4*, *KRT6A*, *KRT8* et *KRT18*), la tropomyosine (*TPM4*), les actines (*ACTB*, *ACTG1*), la thymosine (*TMSB10*) et la « syntenin » (*SDCBP*) ;

2 - les protéines participant à la modification post-traductionnelle des protéines telles la protéine de liaison des lipopolysaccharides (« lipopolysaccharide binding protein », *LBP*) et l'enzyme β1,4-galactosyltransferase (*B4GALT5*) ;

3 - les protéines de la réponse immunitaire, représentées par des protéines du complément (*C3*, *C4B*, *C4BPA*, *CFB*), des interleukines (*IL18BP*, *IL13RA1*, *ILIRN*), un interféron (*IFI30*), une chémokine (*CXCL5*), une intégrine (*ITGB6*), l'annexine 1 (*ANXA1*) et des protéines du complexe majeure d'histocompatibilité (*BoLA-A*) ;

4 - les protéines du métabolisme lipidique parmi lesquelles figurent des protéines de captation des LDL et des chylomicrons avec l'« oxidized low density lipoprotein » (*OLR1*) et l'apolipoprotéine E (*APOE*), des enzymes de dégradation des lipides avec la monoglycéride lipase (*MGLL*) et une lipase (*LIPG*) et enfin une enzyme de synthèse des acyl-CoA longue chaîne avec l'acyl-CoA synthétase chaîne longue (*ACSL5*).

Le gène spécifiant l'enzyme DGAT1 était également significativement sous-exprimé pour le génotype « faible » au locus *DGAT1* (1,3 fois moins exprimé).

Tableau III-6 : Fonctions métaboliques déterminées par « Ingenuity Pathway Analysis » sur les gènes significativement sous-exprimés avec le génotype « faible » (GC/GC, A232) au locus *DGAT1*

Symbole du gène	Protéine codée par le gène	Ln2(R/V)	Ratio d'expression	Numéro de la sonde	Numéro d'accession	Intensités du signal
Gènes sous exprimés pour le génotype GC/GC, A232						**K232/A232**
Signalisation et interaction intercellulaire						
C4BPA	complement component 4 binding protein, alpha	-2,62	6,15	A_73_110994	NM_174252	moyen/faible
C4BPA	complement component 4 binding protein, alpha	-1,40	2,64	Bt00007906	CK950150	faible/faible
CD68	CD68 molecule	-0,78	1,71	A_73_106271	XM_868351	faible/faible
BoLA-A	major histocompatibility complex, class I, A	-0,74	1,67	Bt00007214	CK849467	moyen/faible
LTF	similar to lactotransferrin	-1,54	2,91	A_73_116655	NM_180998	fort/fort
LTF	lactotransferrin	-1,54	2,91	OLIGO_12250	L08604	moyen/moyen
LTF	lactotransferrin	-1,49	2,82	Bt00007015	CK849101	moyen/faible
SBNO2	strawberry notch homolog 2 (Drosophila)	-0,64	1,56	OLIGO_04958	BF043642	faible/faible
Mort cellulaire						
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	-1,18	2,27	A_73_102571	XM_866262	moyen/faible
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	-0,99	1,99	A_73_109426	XM_592737	moyen/moyen
AGPAT2	1-acylglycerol-3-phosphate O-acyltransferase 2 (lysophosphatidic acid acyltransferase, beta)	-0,86	1,81	A_73_118412	XM_589558	moyen/moyen
ALDH1A3	aldehyde dehydrogenase 1 family, member A3	-1,03	2,05	A_73_103806	BE485574	moyen/moyen
ATXN3	ataxin 3	-0,76	1,69	A_73_113215	XM_589821	moyen/faible
ATXN3	ataxin 3	-0,62	1,54	A_73_108214	XM_866678	faible/faible
BIK	BCL2-interacting killer (apoptosis-inducing)	-0,79	1,73	A_73_114020	AW668663	faible/faible
CYBA	cytochrome b-245, alpha polypeptide	-0,63	1,55	A_73_106554	NM_174034	moyen/moyen
DDAH2	dimethylarginine dimethylaminohydrolase 2	-0,59	1,50	A_73_120146	NM_001034704	moyen/moyen
GALNT2	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)	-0,60	1,52	A_73_118965	AV594534	moyen/moyen
HSPA1B	heat shock 70kDa protein 1B	-0,84	1,79	A_73_116780	BP106912	moyen/moyen
RBP1	retinol binding protein 1, cellular	-1,58	2,98	A_73_116738	NM_001025343	moyen/moyen
RBP1	retinol binding protein 1, cellular	-1,22	2,34	OLIGO_10638	DN532441	moyen/faible
SRGN	serglycin	-0,66	1,58	A_73_106384	NM_001025326	moyen/moyen
Mouvement cellulaire						
CAP1	CAP, adenylate cyclase-associated protein 1 (yeast)	-0,64	1,56	A_73_112859	NM_001035010	moyen/moyen
IL18BP	interleukin 18 binding protein	-0,59	1,50	A_73_119324	XM_869739	moyen/moyen
KRT6A	keratin 6A	-0,76	1,70	OLIGO_06876	M14056	moyen/moyen
KRT6A	keratin 6A	-0,70	1,63	A_73_101433	NM_001008663	fort/fort
MGLL	monoglyceride lipase	-0,78	1,72	A_73_104056	AW655967	moyen/moyen
S100A2	S100 calcium binding protein A2	-0,69	1,61	A_73_100329	NM_001034367	faible/faible
SDCBP	syndecan binding protein (syntenin)	-0,84	1,79	A_73_117643	XM_588222	moyen/moyen
SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	-0,76	1,69	A_73_105596	NM_174768	faible/faible
TPM4	tropomyosin 4	-0,92	1,89	A_73_106180	XM_868523	moyen/moyen
Prolifération et croissance cellulaire						
ADRA1B	adrenergic, alpha-1B-, receptor	-0,66	1,58	Bt00001112	CK771135	moyen/faible
AEBP1	AE binding protein 1	-0,62	1,54	A_73_111975	NM_174839	moyen/moyen
AKR1C3	aldo-keto reductase family 1, member C3 (3-alpha hydroxysteroid dehydrogenase, type II)	-2,50	5,65	A_73_100344	NM_181027	moyen/moyen
ARID5B	AT rich interactive domain 5B (MRF1-like)	-0,93	1,90	A_73_113220	CB165266	moyen/moyen
EIF3I	eukaryotic translation initiation factor 3, subunit I	-0,87	1,83	OLIGO_10068	DN642122	moyen/moyen
FER1L3	fer-1-like 3, myoferlin (C. elegans)	-0,71	1,64	A_73_117391	XM_591300	moyen/moyen
IFI30	interferon, gamma-inducible protein 30	-0,81	1,75	A_73_101156	XM_868154	moyen/moyen
KRT4	keratin 4	-0,66	1,58	A_73_100681	XM_880337	faible/faible
OSMR	oncostatin M receptor	-0,76	1,69	A_73_108457	CB434425	moyen/moyen
PDGFC	platelet derived growth factor C	-0,67	1,59	A_73_115702	BE756651	moyen/moyen
PKP1	plakophilin 1 (ectodermal dysplasia/skin fragility syndrome)	-0,85	1,80	A_73_115167	NM_174145	moyen/faible
TACC2	transforming, acidic coiled-coil containing protein 2	-0,62	1,54	A_73_118627	XM_880503	moyen/moyen
TACSTD2	tumor-associated calcium signal transducer 2	-1,64	3,11	A_73_116592	XM_589840	moyen/moyen
ZFP36L1	zinc finger protein 36, C3H type-like 1	-0,76	1,70	A_73_109640	XM_865938	moyen/moyen
ZFP36L2	zinc finger protein 36, C3H type-like 2	-0,98	1,97	A_73_121434	CB419229	moyen/moyen
Signalisation et interaction intercellulaire et mort cellulaire						
KRT18	keratin 18	-1,16	2,23	A_73_118812	XM_582930	moyen/moyen
LBP	lipopolysaccharide binding protein	-1,29	2,45	Bt00006600	CK848266	moyen/faible
LBP	lipopolysaccharide binding protein	-1,15	2,22	A_73_104583	NM_001038674	moyen/moyen
TYROBP	TYRO protein tyrosine kinase binding protein	-0,60	1,52	A_73_118891	NM_174627	moyen/moyen
Signalisation et interaction intercellulaire et mouvement cellulaire						
C4B	complement component 4B (Childo blood group)	-1,06	2,08	A_73_113565	XM_584742	moyen/moyen
C4B	complement component 4B (Childo blood group)	-0,70	1,63	Bt00003648	CK775755	faible/faible
LYPD3	LY6/PLAUR domain containing 3	-0,65	1,57	A_73_116891	BP108674	faible/faible
PLXNC1	plexin C1	-0,66	1,58	A_73_112305	XM_596354	faible/faible
RND3	Rho family GTPase 3	-0,63	1,54	A_73_112008	XM_592424	moyen/moyen
SERPING1	serpin peptidase inhibitor, clade G (C1 inhibitor), member 1, (angioedema, hereditary)	-0,80	1,74	A_73_109408	NM_174821	moyen/moyen
Signalisation et interaction intercellulaire et prolifération et croissance cellulaire						
ACSL5	acyl-CoA synthetase long-chain family member 5	-0,95	1,93	A_73_113054	XM_591964	moyen/moyen
IL13RA1	interleukin 13 receptor, alpha 1	-0,60	1,51	A_73_100295	AW657053	moyen/moyen
LIPG	lipase, endothelial	-0,68	1,60	A_73_100558	XM_591948	faible/faible

RGS2	regulator of G-protein signaling 2, 24kDa	-1,16	2,24	A_73_117206	XM_590677	moyen/moyen
RGS2	regulator of G-protein signaling 2, 24kDa	-1,02	2,03	Bt00007790	CK948092	faible/faible
RGS2	regulator of G-protein signaling 2, 24kDa	-0,73	1,65	OLIGO_09600	CK957537	faible/faible
TACSTD1	tumor-associated calcium signal transducer 1	-0,75	1,68	A_73_108385	NM_001035290	moyen/moyen
Mort et mouvement cellulaire						
B4GALT5	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5	-0,77	1,70	A_73_106983	CB419409	moyen/moyen
CD53	CD53 molecule	-0,64	1,56	A_73_119885	NM_001034232	moyen/moyen
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	-2,86	7,24	A_73_119569	XM_865491	fort/moyen
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	-1,73	3,31	Bt00007310	CK849654	moyen/faible
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)	-1,00	2,00	OLIGO_12367	BT021835	faible/faible
Mort cellulaire et prolifération et croissance cellulaire						
EPHB6	EPH receptor B6	-1,51	2,85	A_73_118831	XM_608259	moyen/moyen
FZD1	frizzled homolog 1 (Drosophila)	-0,60	1,52	A_73_100213	XM_600754	moyen/moyen
IER3	immediate early response 3	-0,93	1,90	A_73_118970	XM_869173	moyen/moyen
IHPK2	inositol hexaphosphate kinase 2	-0,72	1,64	A_73_120783	NM_001034317	moyen/moyen
LDHA	lactate dehydrogenase A	-0,73	1,66	A_73_119375	NM_174099	moyen/moyen
RUNX1	runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)	-0,67	1,59	A_73_112916	CB440904	faible/faible
SCIN	scinderin	-1,08	2,11	A_73_102891	NM_174177	moyen/faible
SCN3B	sodium channel, voltage-gated, type III, beta	-1,31	2,49	A_73_102756	XM_588340	moyen/moyen
TMSB10	thymosin, beta 10	-0,66	1,58	Bt00000870	CK770743	moyen/moyen
TMSB10	thymosin, beta 10	-0,62	1,53	OLIGO_11135	CR550876	moyen/moyen
Mouvement cellulaire et prolifération et croissance cellulaire						
ACTB	actin, beta	-0,79	1,73	A_73_102689	NM_173979	fort/moyen
ACTG1	actin, gamma 1	-0,88	1,84	A_73_114078	NM_001033618	fort/fort
ANKRD1	ankyrin repeat domain 1 (cardiac muscle)	-0,72	1,64	A_73_103476	NM_001034378	faible/faible
BCL3	B-cell CLL/lymphoma 3	-0,63	1,55	A_73_108331	XM_610447	moyen/moyen
BHLHB2	basic helix-loop-helix domain containing, class B, 2	-0,99	1,98	A_73_118295	NM_001024929	moyen/moyen
CFB	complement factor B	-1,72	3,30	A_73_118840	NM_001040526	moyen/moyen
NEURL	neuralized homolog (Drosophila)	-0,68	1,60	A_73_107600	BI849329	moyen/faible
Signalisation et interaction intercellulaire, mort cellulaire et mouvement cellulaire						
OLR1	oxidized low density lipoprotein (lectin-like) receptor 1	-0,99	1,99	A_73_113527	NM_174132	faible/faible
SDC4	syndecan 4	-0,95	1,93	A_73_116016	XM_584869	moyen/moyen
Signalisation et interaction intercellulaire, mort cellulaire et prolifération et croissance cellulaire						
HMOX1	heme oxygenase (decycling) 1	-0,72	1,65	A_73_116548	NM_001014912	moyen/moyen
IL1RN	interleukin 1 receptor antagonist	-0,95	1,93	A_73_101356	NM_174357	moyen/faible
KRT8	keratin 8	-1,27	2,41	A_73_108833	NM_001033610	fort/moyen
KRT8	keratin 8	-1,09	2,12	OLIGO_11338	DN825510	moyen/faible
NFE2	nuclear factor (erythroid-derived 2), 45kDa	-1,06	2,09	A_73_107257	NM_001014923	faible/faible
RASSF5	Ras association (RalGDS/AF-6) domain family 5	-0,67	1,59	A_73_101269	AW632383	moyen/moyen
RELB	v-rel reticuloendotheliosis viral oncogene homolog B, nuclear factor of kappa light polypeptide gene enhancer in B-cells 3 (avian)	-0,61	1,53	A_73_106104	XM_600955	moyen/moyen
Signalisation et interaction intercellulaire, mouvement cellulaire et prolifération et croissance cellulaire						
CLDN11	claudin 11 (oligodendrocyte transmembrane protein)	-0,61	1,52	A_73_118193	NM_001035055	faible/faible
CXCL5	chemokine (C-X-C motif) ligand 5	-1,04	2,05	A_73_106079	NM_174300	faible/faible
GRP	gastrin-releasing peptide	-0,82	1,76	OLIGO_04190	DN287080	faible/faible
MMP19	matrix metallopeptidase 19	-0,82	1,76	A_73_118915	XM_612530	faible/faible
Mort cellulaire, mouvement cellulaire et prolifération et croissance cellulaire						
ANGPTL4	angiopoietin-like 4	-1,72	3,30	A_73_101924	XM_587033	moyen/moyen
ANGPTL4	angiopoietin-like 4	-1,18	2,27	Bt00005718	CK846487	faible/faible
CASP1	caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase)	-0,59	1,50	A_73_121388	XM_592026	faible/faible
DEFB1	defensin, beta 1	-0,67	1,59	A_73_120710	NM_175703	faible/faible
ETV6	ets variant gene 6 (TEL oncogene)	-0,61	1,53	A_73_107505	NM_001015514	moyen/moyen
ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	-0,80	1,74	A_73_101455	CB467551	faible/faible
S100A4	S100 calcium binding protein A4	-0,65	1,57	A_73_118905	NM_174595	moyen/moyen
SOD2	superoxide dismutase 2, mitochondrial	-0,87	1,82	A_73_102060	NM_201527	moyen/moyen
Signalisation et interaction intercellulaire, mort cellulaire, mouvement cellulaire et prolifération et croissance cellulaire						
ANXA1	annexin A1	-0,72	1,64	A_73_114465	NM_175784	moyen/faible
ANXA1	annexin A1	-0,67	1,59	Bt00007197	CK849435	faible/faible
APOE	apolipoprotein E	-0,74	1,68	A_73_100548	NM_173991	moyen/moyen
C3	complement component 3	-0,61	1,53	OLIGO_06959	CB461274	moyen/faible
CLU	clusterin	-1,79	3,46	A_73_120760	NM_173902	fort/moyen
CLU	clusterin	-1,51	2,84	OLIGO_03474	CN441848	moyen/moyen
CLU	clusterin	-0,99	1,99	Bt00006868	CK848824	faible/faible
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa	-0,63	1,55	OLIGO_11225	NM_001004024	faible/faible
DDR1	discoidin domain receptor family, member 1	-0,88	1,83	A_73_113008	XM_879506	moyen/moyen
DKK3	dickkopf homolog 3 (Xenopus laevis)	-0,62	1,53	A_73_116904	XM_583102	moyen/moyen
EFNA1	ephrin-A1	-0,83	1,78	A_73_112586	NM_001034292	moyen/moyen
EGR1	early growth response 1	-1,49	2,81	A_73_111154	XM_601394	moyen/moyen
ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	-0,88	1,84	A_73_115898	XM_585288	moyen/moyen
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	-0,66	1,58	A_73_100116	XM_870743	moyen/moyen
ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	-1,26	2,40	A_73_120913	BP107912	fort/moyen
FHL2	four and a half LIM domains 2	-0,59	1,50	A_73_113241	XM_878142	moyen/moyen
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	-1,71	3,27	A_73_101038	NM_182786	moyen/moyen

FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	-0,87	1,82	OLIGO_09217	CV983373	faible/faible
ITGB6	integrin, beta 6	-1,70	3,24	A_73_104506	NM_174698	moyen/faible
MET	met proto-oncogene (hepatocyte growth factor receptor)	-0,91	1,88	A_73_116287	NM_001012999	moyen/moyen
MET	met proto-oncogene (hepatocyte growth factor receptor)	-0,69	1,61	Bt00005375	CK845891	faible/faible
MSR1	macrophage scavenger receptor 1	-0,64	1,55	A_73_102799	NM_174113	moyen/faible
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	-0,90	1,86	A_73_112809	CB464624	moyen/moyen
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (p85 alpha)	-0,59	1,50	Bt00003677	CK775806	faible/faible
PTGES	prostaglandin E synthase	-0,94	1,92	A_73_111837	NM_174443	moyen/moyen
S100A10	S100 calcium binding protein A10	-0,59	1,51	A_73_104422	NM_174650	moyen/moyen
S100A9	S100 calcium binding protein A9	-0,93	1,90	A_73_110650	XM_611685	faible/faible
ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	-0,65	1,57	A_73_110203	U30854	moyen/moyen
TGFBI	transforming growth factor, beta-induced, 68kDa	-0,60	1,52	A_73_112587	XM_588272	moyen/faible
THBS2	thrombospondin 2	-0,86	1,82	A_73_102889	NM_176872	moyen/faible
TMSB4X	S100 calcium binding protein G	-0,95	1,94	A_73_106202	NM_001002885	fort/fort
TNC	tenascin C (hexabrachion)	-0,80	1,75	A_73_108252	XM_612563	faible/faible
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B	-0,69	1,61	A_73_113320	NM_001040490	moyen/moyen
VCAM1	vascular cell adhesion molecule 1	-0,60	1,51	A_73_118327	NM_174484	moyen/moyen

➢ Résultats des quantifications en PCRq

Afin de valider les données observées avec la technique des puces à ADN, les transcrits des gènes spécifiant DGAT1, les caséines α_{s1} et κ (choisis arbitrairement) ont été quantifiés par PCRq. Les ratios entre les deux génotypes au locus *DGAT1* ont été calculés sur les quantités de transcrits normalisées. Ces ratios obtenus en PCRq ont été comparés avec ceux obtenus avec la technique des lames de verre (cf. Figure III.5). La PCRq a confirmé les résultats obtenus sur microarrays. On note pour ces trois gènes une bonne concordance entre les techniques. L'allèle « faible » au locus *DGAT1* est associé à une quantité de transcrits plus faible, alors que les messagers issus des gènes spécifiant les caséines α_{s1} et κ sont au contraire en plus grand nombre (ratios génotype « faible »/génotype « fort » de 0,6 ; 1,5 et 1,3 respectivement).

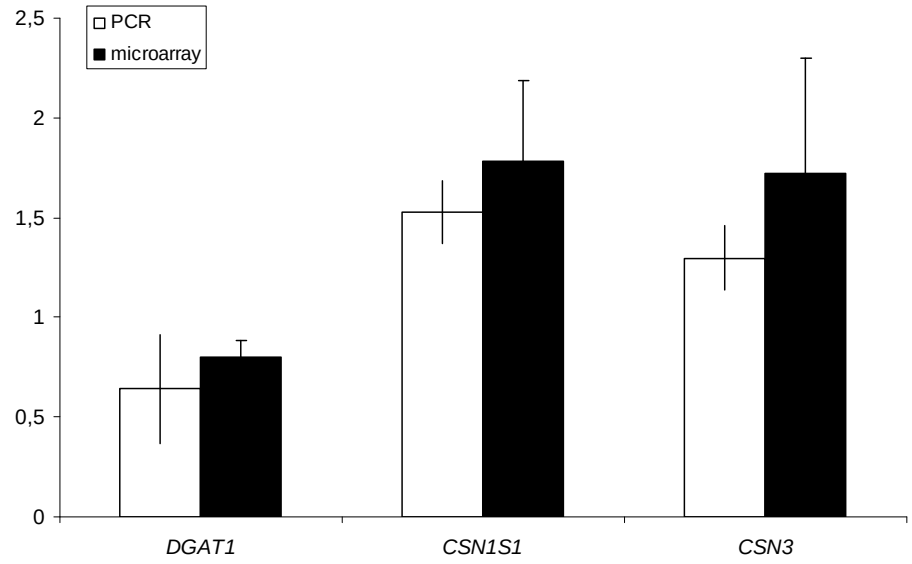


Figure III-5 : Comparaison des quantifications en PCR quantitative (histogramme blanc) et en microarray (histogramme noir) exprimées en moyenne des ratios de transcrits DGAT1 entre génotypes (GC, A232 / AA, K232). Les barres verticales représentent l'écart-type.

➢ Relation entre gènes différemment exprimés et données phénotypiques

Pour mettre en relation les gènes différemment exprimés entre les deux génotypes au locus *DGAT1* avec les données phénotypiques, une ACP a été réalisée sur l'intensité moyenne du signal des 538 sondes différemment exprimées avec un seuil de significativité de 10% et une limite sur rapport d'expression de 1,5 fois. La première dimension de l'ACP explique 53% de la variabilité et discrimine les deux génotypes. 533 sondes, 270 avec des coordonnées négatives et 263 avec des coordonnées positives caractérisent la première dimension. Les individus de génotype « faible » se distinguent par des coordonnées négatives sur cette dimension, alors que les individus de génotype « fort » ont des coordonnées positives. Les coordonnées des individus sur la première dimension ont été importées en qualité de variable qualitative pour une analyse de variance effectuée sur les données phénotypiques.

Les teneurs en AG C8:0, C10:0 et C14:0 étaient significativement plus élevées alors que les teneurs en AG C14:1, C16:1 *cis9* et C17:1 étaient significativement plus faibles avec des coordonnées négatives, associées dans cette analyse au génotype « faible » (cf. Tableau III.7). Les index de désaturation des AG C10:0, C14:0 et C16:0 étaient significativement plus faibles avec des coordonnées négatives, de même pour la taille des globules gras. En

revanche, les teneurs en caséines α_{s1} et β étaient significativement plus élevées avec des coordonnées négatives, et donc avec le génotype « faible ».

Tableau III-7 : Résultats sur la mise en relation des données transcriptomiques et phénotypiques. Seules les variables phénotypiques significativement différentes entre les coordonnées négatives et positives de l'ACP sont présentées dans ce tableau. Ne sont indiquées ici que les 15 première sondes qui caractérisent le plus chaque direction de la première dimension sont indiquées dans ce tableau.

	Dimension 1 négative (« Faible », GC/GC, A232)	**Dimension 1 positive (« Fort », AA/AA, K232)**	
Variables phénotypiques			
	moyenne (ecart-type)	moyenne (ecart-type)	Pc
C8:0 (%)	1,57 (0,13)	1,48 (0,14)	P<0,1
C10:0 (%)	4,09 (0,29)	3,73 (0,61)	P<0,1
C14:0 (%)	13,52 (0,21)	12,66 (1,24)	P<0,05
C14:1 (%)	1,05 (0,13)	1,16 (0,09)	P<0,1
C16:1*cis*9 (%)	1,26 (0,19)	1,50 (0,24)	P<0,05
C17:1 (%)	0,17 (0,02)	0,19 (0,01)	P<0,05
s10et12et14	22,77 (0,84)	21,25 (2,69)	P<0,1
index10**	8,01 (0,35)	9,10 (1,23)	P<0,05
index14**	7,12 (0,79)	8,37 (1,14)	P<0,05
index16**	3,66 (0,35)	4,15 (0,67)	P<0,05
TGG (µm)	4,52 (0,44)	4,76 (0,57)	P<0,05
Teneur caséine α_{S1} (%)	28,14 (1,45)	27,04 (1,80)	P<0,05
Teneur caséine β (%)	34,98 (1,24)	34,05 (1,48)	P<0,1
Top 15 des sondes expliquant le plus la 1ère dimension			
	CIDEA (cell death-inducing DFFA-like effector a)	GALC (galactosylceramidase)	
	CIDEA (cell death-inducing DFFA-like effector a)	PIK3IP1 (phosphoinositide-3-kinase interacting protein 1)	
	ATP13A4 (ATPase type 13A4)	DDR1 (discoidin domain receptor family, member 1)	
	FAM46C (family with sequence similarity 46, member C)	FZD7 (frizzled homolog 7 (Drosophila))	
	DNAJC12 (DnaJ (Hsp40) homolog, subfamily C, member 12)	MET (met proto-oncogene (hepatocyte growth factor receptor))	
	FABP3 (fatty acid binding protein 3, muscle and heart)	OLR1 (oxidized low density lipoprotein (lectin-like) receptor 1)	
	AGPAT6 (1-acylglycerol-3-phosphate O-acyltransferase 6)	KRT8 (keratin 8)	
	BRP44L (brain protein 44-like)	IHPK2 (inositol hexaphosphate kinase 2)	
	ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2)	RIN2 (Ras and Rab interactor 2)	
	BDH1 (3-hydroxybutyrate dehydrogenase, type 1)	ITGB6 (Integrin B6)	
	MCFD2 (multiple coagulation factor deficiency 2)	BIK (BCL2-interacting killer (apoptosis-inducing))	
	TMEM64 (transmembrane protein 64)	SASH1 (SAM and SH3 domain containing 1)	
	NUCB2 (nucleobindin 2)	NFE2 (nuclear factor (erythroid-derived 2), 45kDa)	
	RUNDC1 (RUN domain containing 1)	TACSTD2 (tumor-associated calcium signal transducer 2)	
	MYBPC1 (Myosine Binding protein)	ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3)	
	SLC34A2 (solute carrier family 34 (sodium phosphate), member 2)	LTF (lactotransferrin)	

* ACP = Analyse en Composante Principale ; Pc = probabilité critique ; TGG = taille globules gras ; s10et12et14 = C10:0 (%) + C12:0 (%) + C14:0 (%) ; ** les index de désaturation sont calculés comme indiqué pour l'exemple suivant : index14 = C14:1 *cis9*/(C14:1 *cis9*+ C14:0)

Comme la totalité des sondes entrées dans l'analyse ACP discriminait les individus selon leur génotype au locus *DGAT1*, nous avons choisi de présenter uniquement les quinze premières sondes les plus discriminantes.

Les coordonnées négatives de l'axe 1 se caractérisaient par des gènes représentatifs du métabolisme lipidique (*FABP3*, *AGPAT6*, *ABCG2*, *BDH1*), du transport intra-cellulaire (*ATP13A4*, *DNAJC12*, *MCFD2*, *NUCB2*) et de l'apoptose (*CIDEA*).

Les gènes les plus explicatifs des coordonnées positives de l'axe 1 spécifient des protéines qui participent à l'adhésion cellulaire (*DDR1*, *OLR1*, *ITGB6*), l'apoptose (*IHPK2*, *BIK*), la transduction du signal (*FZD7*, *MET*, *RIN2*, *TACSTD2*), l'organisation cellulaire (*KRT8*), la régulation de la transcription (*NFE2*) et à la régulation négative du cycle cellulaire (*SASH1*) et de l'apoptose (*ERBB3*).

III.3.4. Discussion

III.3.4.1. Points forts et points faibles du protocole expérimental

Cette étude est la première du genre, conduite chez la vache laitière, portant sur l'impact du polymorphisme au locus *DGAT1* sur des caractères laitiers et sur l'expression de plusieurs milliers de gènes dans la glande mammaire, et utilisant des couples de vaches pleine ou demi-sœurs de génotypes homozygotes opposés à ce locus. Lors des prélèvements, les animaux étaient conduits dans un milieu extrêmement bien contrôlé, sur une même installation expérimentale. Ils recevaient le même régime alimentaire et étaient au même nombre de jours de lactation.

L'hybridation des cibles marquées a été effectuée sur deux types de lame. L'utilisation combinée de ces deux supports a permis d'augmenter le nombre de gènes étudiés par comparaison à l'utilisation d'un seul type de lame. Les 14 567 gènes (avec une annotation IPA) distincts présents sur les deux lames cumulées représentent environ 55% du génome bovin. La concaténation des 2 fichiers des deux types de lames a permis, après validation de la procédure (vérification de la cohérence des jeux de données avant et après concaténation), de traiter toutes les sondes de la même façon sans tenir compte de leur origine, en appliquant le test multiple sur l'ensemble des données disponibles. A ce jour, nous n'avons pas trouvé d'études équivalentes décrites dans la littérature ayant procédé de la sorte. De plus, la redondance de certains gènes entre les deux types de lame (taux de recouvrement de 25%) a permis de valider les différentiels d'expression de ces gènes lorsqu'ils étaient déclarés différemment exprimés entre les deux génotypes. En effet, lorsque les intensités d'une sonde étaient significativement différentes entre les deux génotypes, cette différence se retrouvait pour une autre sonde du même gène, avec le même sens de variation.

Par ailleurs, le différentiel d'expression de trois gènes (*CSN1S1*, *CNS3* et *DGAT1*) a été validé par trois systèmes de quantification en PCRq. La validation du différentiel d'expression de quelques autres gènes permettra de valider, pour partie, l'expérience des puces à ADN réalisée dans cette étude.

L'impossibilité de dissocier les effets des polymorphismes des lactoprotéines majeures de l'effet de la mutation K232A au locus *DGAT1*, pour l'analyse de la quantité des protéines, montre une des limites de notre étude qui est le faible nombre d'animaux du dispositif. En

effet, certains variants protéiques sont uniquement présents pour l'un des génotypes au locus *DGAT1*, comme par exemple le variant AA de la caséine κ qui n'a été détecté qu'une seule fois sur les 11 laits individuels analysés, chez un individu de génotype « faible ». Or, le polymorphisme génétique des protéines peut impacter le niveau d'expression allélique et donc la quantité de protéine (pour revue Martin *et al.*, 2002). Nous ne pouvions donc pas conclure quant à l'existence d'un lien entre le polymorphisme au locus *DGAT1* et la quantité de caséine κ. Pour conclure, il faudrait augmenter l'effectif des animaux en veillant à croiser et équilibrer les génotypes au locus des protéines majeures et le génotype au locus *DGAT1*.

Dans cette expérience, nous avons essayé d'associer, par des méthodes statistiques, une liste de gènes déclarés différemment exprimés avec des données phénotypiques. Pour cela, nous avons appliqué une ACP sur les gènes dont l'expression est affectée, puis nous avons effectué une analyse de variance (ANOVA) sur les données phénotypiques, à partir des coordonnées des individus sur l'axe de l'ACP qui discriminait le plus le génotype au locus *DGAT1*. Cette méthode est hautement critiquable puisqu'elle reste très dépendante du choix de la personne qui analyse les données, et ne nous a pas permis d'extraire un sens biologique à nos observations. De nouvelles méthodes d'analyse de biologie intégrative sont en cours de construction dans de nombreux laboratoires de statistique (à la Station d'Amélioration Génétique Animale, SAGA, INRA de Toulouse, par exemple) et pourront être appliquées à l'ensemble des données produites dans cette expérience.

III.3.4.2. Le dysfonctionnement de l'enzyme DGAT1 par la mutation K232A implique d'autre voie de synthèse pour les lipides du lait

Les résultats montrent que le génotype « faible » au locus *DGAT1* est corrélé à une composition de la matière grasse laitière plus riche en AG longs et insaturés, ainsi que des index de désaturations (caractérisant l'activité de la stéaroyl CoA désaturase 1) plus faibles pour les acides C10:0, C14:0 et C16:0. Ces observations avaient déjà été faites par Schennink et al. (Schennink *et al.*, 2008; Schennink *et al.*, 2007) qui supposaient alors que la variation de la composition de la matière grasse laitière liée au polymorphisme K232A au locus *DGAT1* était induit par un changement de spécificité de substrat de l'enzyme consécutivement à la mutation qui transforme un résidu Lysyle en résidu Alanyle. Or, la modification de la

composition en AG du lait s'accompagne surtout d'une diminution significative du TB pour le génotype « faible » qui serait imputable, selon Grisart *et al.*, (2004), à une perte d'efficacité de l'enzyme DGAT1. Cette enzyme catalysant l'estérification d'un acyl-CoA en position 3 d'une molécule de diacylglycérol, une diminution de l'efficacité de l'enzyme devrait s'accompagner d'une accumulation de diacylglycérols dans la cellule épithéliale mammaire.

Dans les cellules, il existe deux voies possibles d'utilisation du diacylglycérol. La première concerne des voies de synthèse des triglycérides alternatives à DGAT1 faisant intervenir soit une enzyme de transfert d'AG entre deux molécules de diacylglycérol, la sn-1,2(2,3)-diacylglycérol transacylase (Lehner & Kuksis, 1993), soit une seconde enzyme diacylglycérol acyltransferase, la DGAT2 (Cases *et al.*, 2001) qui possède la même activité que l'enzyme DGAT1. La deuxième voie métabolique utilisant le diacylglycérol est la synthèse des phospholipides membranaires faisant intervenir différentes enzymes souvent concurrentes de l'enzyme DGAT1 (Bell *et al.*, 1981). Toutes ces voies alternatives peuvent expliquer la modification de composition en AG observée entre les deux génotypes *DGAT1*.

En effet, l'enzyme sn-1,2(2,3)-diacylglycérol transacylase transfert un AG d'un diacylglycérol vers une autre molécule de diacylglycérol, générant un triacylglycérol et un monoacylglycérol (Lehner & Kuksis, 1993). Dans le lait, les positions sn-1 et sn-2 des triacylglycérols sont préférentiellement estérifiées par le C16:0, le C14:0, le C18:1 et le C18:0 (Breckenridge & Kuksis, 1968), ce qui expliquerait, pour partie, l'augmentation significative de la teneur du lait en C14:0 avec le génotype « faible » au locus *DGAT1*.

Dans notre étude, le gène spécifiant l'enzyme DGAT2 était deux fois plus exprimé avec le génotype *DGAT1* « faible » (donnée non montrée, P=0,11). L'activité de l'enzyme DGAT2 purifiée à été étudiée uniquement chez *Mortierella ramanniana* (un champignon) et est plus importante pour les acyl-CoA à chaîne moyenne (C12:0) et les diacylglycérols à chaîne courte et moyenne (C6:0, C8:0 et C10:0) (Lardizabal *et al.*, 2001). Si l'on s'accorde une comparaison entre l'activité de DGAT2 de *Mortierella ramanniana* et celle du bovin, la spécificité d'activité de DGAT2 expliquerait pour partie l'élévation significative de la teneur du lait en AG en C10 et C14 avec le génotype « faible ». Il a été montré sur des cellules HeLa une colocalisation entre les enzymes Stearoyl CoA désaturase 1 (SCD1) et DGAT2 (Man *et al.*, 2006). On peut faire l'hypothèse que cette colocalisation puisse également exister dans le tissu mammaire et que les acylCoA désaturés par la SCD1 servent directement de substrat à l'enzyme DGAT2, expliquant ainsi l'augmentation des AG à chaîne longue et insaturée avec le génotype *DGAT1* « faible ».

Ce génotype se caractérise également par des globules gras plus petits. Or, les globules gras sont des structures composées de triglycérides enrobés de deux couches membranaires (pour revue Couvreur & Hurtaud, 2007). A quantité égale de matière grasse, plus les globules gras sont petits, plus la surface membranaire est importante. Or nous n'avons pas observé de différence en terme de quantité de matière grasse sécrétée entre les deux génotypes *DGAT1*, ce qui suggère qu'il puisse être nécessaire de produire davantage de membrane avec le génotype « faible ». Avec ce génotype, les gènes spécifiant les enzymes phosphatidylcholine-retinol O-acyltransférase (LRAT) et phosphatidylethanolamine N-methyltransférase (PEMT) qui participent à la synthèse des phospholipides membranaires (Bell *et al.*, 1981), sont significativement surexprimés. Les diacylglycérols accumulés dans le tissu des animaux de génotype « faible » pourraient servir à la synthèse de phospholipides membranaires nécessaires à la sécrétion des globules gras de plus petite taille.

Au-delà d'une modification de la spécificité de substrat de l'enzyme DGAT1 par la mutation K232A (Schennink *et al.*, 2007), le changement significatif de composition en AG du lait s'expliquerait par l'emprunt de voies alternatives du métabolisme du diacylglycérol dans le tissu mammaire, telles que la synthèse des triglycérides par d'autres enzymes que DGAT1 et la synthèse des phospholipides membranaires qui constituent environ 3% des lipides du lait. On s'attendrait avec le génotype « faible » à une augmentation de la teneur du lait en phospholipides, hypothèse que nous n'avons pas eu la possibilité de vérifier.

III.3.4.3. La mutation K232A au locus DGAT1 modifie l'activité métabolique du tissu mammaire par une augmentation de sensibilité du tissu à l'insuline

L'analyse transcriptomique révèle une augmentation de l'expression des gènes participant au métabolisme des glucides et à la synthèse des lipides du lait avec le génotype « faible » au locus *DGAT1*. De plus, l'ACP sur les gènes différemment exprimés entre génotypes montre que le gène spécifiant le CIDEA discrimine le plus les animaux selon le génotype au locus *DGAT1*. Ce gène est surexprimé avec le génotype « faible ». Chez des souris obèses insulino-résistantes, il y a une diminution de 50% des transcrits provenant du gène *CIDEA*, par comparaison aux souris non-obèses (Nordstrom *et al.*, 2005). La perte d'efficacité (Grisart *et al.*, 2004) et le changement de spécificité de substrat (Schennink *et al.*,

2007) de l'enzyme DGAT1 induit par la mutation K232A, augmenterait la sensibilité des tissus mammaires à l'insuline chez les individus de génotype « faible » comme cela a été montré chez des souris déficientes en enzyme DGAT1 (Chen *et al.*, 2002).

Cette hypothèse se vérifie par la surexpression, avec le génotype « faible », de plusieurs gènes dont l'expression dépend de l'insuline tels que les gènes spécifiant DGAT2 (Meegalla *et al.*, 2002) et l'adiponectine (Chen *et al.*, 2003).

La surexpression de l'adiponectine chez les individus de génotype faible apparaîtrait comme un élément clé dans le mécanisme d'augmentation de la sensibilité des tissus mammaires à l'insuline. Chez des souris normales greffées avec du tissu adipeux de souris invalidées pour le gène *DGAT1*, une augmentation de la sensibilité à l'insuline due à la sécrétion de facteurs tels que l'adiponectine par les adipocytes greffés, a été observée (Chen *et al.*, 2003). Cela signifie donc que la présence d'adiponectine en quantité plus importante dans les tissus mammaires des vaches de génotype « faible » pourrait expliquer la sensibilité accrue de ce tissu à l'insuline.

Par ailleurs, il a été montré chez la souris que l'augmentation de la teneur en adiponectine du sérum est corrélée à une plus grande mobilisation des réserves corporelles (Yamauchi *et al.*, 2001). Ceci pourrait être valable aussi chez les vaches de génotype « faible » pour lesquelles la teneur des AG longs et insaturés augmente dans le lait. Ces AG auraient pour origine la mobilisation des réserves corporelles de l'animal.

De plus, les tissus mammaires des individus de génotype « faible » sur-expriment des gènes de la synthèse du lactose (*GK, LALBA*), de la captation, de la synthèse, du transport et de la sécrétion des lipides du lait (*CD36, LPL, ACACA, FABP3, FABP4, GPAM, AGPAT6, ADFP, BTN1A1, ABCG2*) et de la synthèse et de la sécrétion des protéines du lait (*CSN1S1, CSN1S2, LALBA*). Les souris invalidées pour le gène *DGAT1* montrent une augmentation de la dépense énergétique et du métabolisme glucidique (Smith *et al.*, 2000). Dans le tissu mammaire, le métabolisme glucidique serait également affecté et détourné vers la synthèse de lactose, en réponse à une sensibilité accrue des tissus à l'insuline, expliquant l'augmentation de la quantité de lactose et donc de lait sécrété chez les individus de génotype « faible ». Schennink (communication personnelle) confirme un effet de la mutation K232A au locus *DGAT1* sur la quantité de lactose sécrétée, dans le sens d'une augmentation avec le génotype « faible ».

Les individus de génotype « faible » présenteraient une sensibilité accrue à l'insuline qui se manifesterait par la surexpression des gènes spécifiant le CIDEA,

l'adiponectine et les protéines intervenant dans la synthèse et la sécrétion du lactose, des lipides et des protéines du lait.

III.3.5. Conclusion

La mutation K232A au locus *DGAT1* situé sur le chromosome BTA14 entraîne une modification de composition en AG des laits vers une teneur plus élevée en AG longs et insaturés chez les individus de génotype « faible » (GC/GC, A232). Cette modification aurait pour effet d'induire une réorientation des voies de synthèse des lipides du lait (triglycérides et lipides membranaires). La tendance à l'augmentation des quantités de lait et de lactose sécrétées observée chez les individus de génotype « faible » s'expliquerait par une surexpression des gènes spécifiques de la voie de synthèse du lactose, représentative d'une sensibilité accrue des tissus mammaires à l'insuline. Cette hypothèse reste à valider en opérant des analyses complémentaires sur ces tissus, tant au niveau transcriptionnel, en mesurant par PCRq l'expression de gènes sensibles à l'insuline (*CIDEA*, *DGAT2*, *ADIPOQ*), qu'au niveau protéique, en comparant les deux génotypes *DGAT1* par électrophorèse bidimensionnelle différentielle multiplexée (2D-DIGE).

Les limites de cette expérience résident dans le faible effectif en animaux du dispositif, ce qui a rendu impossible l'observation d'un lien entre la mutation K232A et la quantité des différentes protéines majeures du lait, alors qu'un rapport avait été établi entre le TP et le QTL au BTA14 (Boichard *et al.*, 2003). Par ailleurs, la mise en relation entre les données phénotypiques et les profils d'expression des gènes s'est avérée peu concluante puisque la discrimination des individus selon leur génotype s'expliquait par tous les gènes entrés dans le système, alors que nous attendions une liste plus réduite au seuil fixé. Il serait intéressant de trouver d'autres méthodes pour l'intégration de ce type de données telles que la méthode de sélection de données (« Partial Least Square regression », PLS) développées par Lê Cao (Lê Cao *et al.*, 2008).

Discussion générale et conclusions

L'émergence de nouvelles technologies à haut-débit et du concept d'analyse globale (ou intégrée) pour progresser dans la compréhension du fonctionnement des tissus mammaires (approche transcriptomique et protéomique) et de ses effets sur la composition fine du lait (AG et protéines), modifie assez profondément la manière d'aborder l'étude de l'organisation et du fonctionnement des organismes et donc de concevoir les actions de recherche. En effet, les approches exploratoires, basées sur l'utilisation des disciplines en « omiques », génère un ensemble de données qui une fois analysé, permet de poser de nouvelles hypothèses quant à la compréhension des mécanismes biologiques qui sous-tendent, aux niveaux moléculaires et cellulaires, la synthèse et la sécrétion du lait.

Par ailleurs, la demande sociétale, avec l'apparition d'un consommateur de plus en plus soucieux de sa santé et de l'origine des produits qu'il achète et consomme, change la façon de penser et de gérer les filières. Dans la filière laitière, il devient indispensable de comprendre les mécanismes de mise en place et de fonctionnement de la glande mammaire, afin de déterminer les leviers qui permettront de modifier et/ou d'adapter la composition du lait.

Les travaux auxquels j'ai participé, dans le cadre de ma thèse, prennent en compte ce contexte et ces nouveaux enjeux, qu'il s'agisse du programme relatif à la différentiation terminale des tissus mammaires caprins au cours de la gestation (manuscrit 1) ou du programme « Genomilk Fat », financé par l'ANR Genanimal et les partenaires professionnels (Apis-gène), dont l'objectif est d'étudier l'effet de polymorphismes génétiques (locus *CSN1S1* chez la chèvre et locus *DGAT1* chez la vache), de l'alimentation et de l'interaction entre ces deux facteurs, sur le fonctionnement de la glande mammaire et sur la composition et les caractéristiques des structures supramoléculaires du lait (globules gras et micelles de caséines).

Ces travaux de thèse se caractérisent par la grande diversité du matériel biologique disponible et mis en oeuvre. En effet, nous avons pu, dans le cadre du projet caprin, étudier 5 stades de développement dont le début de la gestation ; et dans le cadre du programme « Genomilk Fat », comparer deux génotypes opposés (homozygotes pour deux allèles

différents) au locus *DGAT1*, responsables de phénotypes contrastés, en matière de TB et de profils en AG.

Pour obtenir le maximum d'information sur ce matériel biologique, nous nous sommes attaché à produire, avec les outils disponibles alors, des profils d'expression génique au moyen de puces à ADN. Les résultats majeurs issus de ces travaux de thèse sont avant tout la production de 2 ensembles de données, dont l'analyse, partielle, a permis une avancée significative dans la compréhension des mécanismes biologiques qui se mettent en place dans la glande mammaire pendant la gestation, chez la chèvre et des voies métabolique qui sont affectés par le polymorphisme au locus *DGAT1*, chez la vache. En second lieu, les travaux réalisés dans le cadre de cette thèse m'ont permis d'amorcer une réflexion sur la technologie des puces à ADN.

Réflexion sur la technologie des puces à ADN

Avec l'évolution des technologies et le développement d'outils nouveaux, nous avons été amenés à utiliser 3 types de puces à ADN dont les sondes (oligonucléotides) ont été déposées (spottées) ou synthétisées *in situ*. Par ailleurs, le matériel et les équipements mis à disposition sur la plateforme PiCT de l'Inra de Jouy-en-Josas, où les expériences ont été réalisées, a également évolué, ce qui m'a permis de tester différents protocoles de marquage, d'hybridation, différentes procédures d'acquisition d'image et de quantification des signaux.

Cette expérience m'a permis de mieux cerner les difficultés et les pièges de la procédure expérimentale, de réaliser que cette technologie était encore insuffisamment stabilisée et devait être adaptée, à chaque nouvelle expérience, au type et au lot de lames utilisés. Cela a nécessité beaucoup d'effort et d'abnégation et une grande partie du travail a été consacrée à l'adaptation des protocoles expérimentaux au type de support et aux équipements disponibles.

Ainsi, ma première expérience de transcriptomique qui avait pour objet la comparaison « début de gestation » *vs*. « lactation » (G46-L40) a été réalisée sur des lames, « spottées » sur la plateforme Biopuces de la Génopole de Toulouse-Midi Pyrénées et sur lesquelles avait été déposés un set de 8 329 oligonucléotides synthétisés par la société Operon complété par une série de 50 oligo représentatifs d'autant de gènes exprimés dans le tissu mammaire (numéro d'accession GEO GPL4594, (Ollier *et al.*, 2007). Ce premier support a permis de comparer l'expression d'environ 6 656 gènes différents, ayant une annotation acceptée par le logiciel d'interprétation biologique, IPA.

L'étude de la différentiation terminale du tissu mammaire caprin sur 5 stades de développement s'est appuyée sur des lames 22K bovines, produites au CRB-GADIE de l'Inra de Jouy-en-Josas, sur lesquelles avaient été spottés deux sets de sondes oligonucléotidiques « Operon » (8 329) et « Illumina » (13 303). Ce type de support, considérablement enrichi en sondes pour atteindre un répertoire de plus de 21 600 éléments, par défaut d'annotation ne permettait la comparaison des tissus que sur environ 10 000 gènes différents ayant une annotation acceptée par IPA (9 973 précisément, soit 38% du génome bovin).

Pour analyser l'effet de la mutation K232A au locus *DGAT1*, sur l'activité transcriptionnelle du tissu mammaire bovin, nous avons utilisé simultanément les lames 22K bovines du CRB et les lames 4 x 44 K bovines commercialisées par la société Agilent (Agilent Technologies). La combinaison des répertoires de gènes représentés sur ces deux lames revenait à étudier l'expression d'environ 14 567 gènes différents ayant une annotation acceptée par IPA, soit environ 55% du génome bovin.

La diversité du matériel expérimental utilisé ouvre quatre pistes de réflexion relativement à la technologie des puces à ADN deux couleurs pour l'étude du transcriptome.

a) l'une des études conduites s'appuie sur des hybridations en système hétérologue, c'est-à-dire que la cible complexe et les sondes spottées provenaient de deux espèces différentes (chèvre sur bovin). Cette pratique est classiquement utilisée lorsque peu de données génomiques sont disponibles pour l'espèce étudiée et lorsque les deux espèces sont phylogénétiquement proches et présentent une forte similarité de séquence (Bar-Or *et al.*, 2007). Actuellement, les programmes de séquençage des génomes n'incluent pas encore le génome caprin et le faible nombre de séquences caprines disponibles dans les bases de données, nous a conduit à utiliser des puces bovines. La validité de l'approche hétérologue : hybridation d'ARN caprin sur lames bovines a été démontrée par (Ollier, 2007), dans le cadre de sa thèse (Université d'Auvergne, Clermont-Ferrand I) où l'hybridation croisée de tissus mammaires caprin et bovin en lactation sur des lames bovines (Operon) a montré une corrélation de 93% entre les tissus des deux espèces.

b) L'utilisation de différentes puces à ADN et de différents matériels de laboratoire au cours de ces trois années de thèse est le reflet d'une évolution et d'une amélioration de la technologie des puces à ADN. L'hybridation des lames « Operon » en chambre manuelle a impliqué une grande intervention du manipulateur alors que le protocole expérimental utilisé avec les lames Agilent était très automatisé, contrôlé et reproductible. Le point central des puces à ADN est donc une homogénéité du traitement des lames, tant sur le protocole

expérimental, que sur l'acquisition et la quantification des images, dans le but d'obtenir des résultats reproductibles d'un manipulateur à un autre, et de réduire au maximum les biais liés à la manipulation. Nous avons systématiquement observé une meilleure comparaison des lames entre elles avec les lames à sondes synthétisées *in situ* (lames Agilent). La vérification de la reproductibilité des résultats avec ces lames permettrait de réduire les coûts expérimentaux en autorisant l'hybridation d'une seule condition par lame.

c) L'augmentation du nombre de gènes étudiés en une seule fois a été permise par la combinaison de différents sets oligonucléotidiques sur une lame et la combinaison de plusieurs lames dans une étude. Ainsi, le nombre de gènes différents analysés, ayant une annotation acceptée par IPA, est passé de 6 656 gènes avec la lame « Operon » à 14 567 gènes en combinant les lames « CRB » et « Agilent ». Cette observation souligne la marge de progression existante et la possibilité d'améliorer l'information d'une puce d'expression par la combinaison de différents répertoires.

d) L'utilisation de différents logiciels pour interpréter biologiquement les données générées par la technique des puces à ADN requiert l'obtention d'annotations adaptées. Or, l'annotation des sondes présentes sur les lames se fait à partir des séquences oligonucléotiques 70-mer (lames « Opéron » ou « CRB ») ou 60-mer (lames Agilent). Il est donc parfois impossible d'obtenir une annotation précise de la sonde, en terme de gène, si le contig auquel appartient la séquence n'a pas encore été associé à un gène. C'est pour cette raison que le nombre de gènes ayant une annotation acceptée par le logiciel d'interprétation biologique IPA, est souvent bien inférieur au nombre de sondes différentes présentes sur la lame. Ainsi, une amélioration de l'annotation des gènes exprimés dans le tissu mammaire permettra de perfectionner l'interprétation biologique des données produites dans le cadre de cette thèse. Le projet d'annotation des gènes exprimés dans différents états physiologiques du tissu mammaire est l'un des objectifs de l'International Milk Genomics Consortium initié par la « California Dairy Research Foundation » et l'Université de Californie Davis (http://lactoknow.ucdavis.edu/), auquel contribuent, pour la France, l'Inra et le Centre National Interprofessionnel de l'Economie Laitière (CNIEL).

Sur le plan conceptuel, les deux projets qui m'ont été confiés m'ont permis de découvrir la construction de plusieurs plans expérimentaux et les différentes analyses possibles des données issues d'une expérience mettant en jeu la technologie des puces à ADN. Deux types de plans expérimentaux ont été construits, l'un en boucle permettant d'effectuer une analyse cinétique pour suivre les différentes étapes du processus de différenciation

terminale du tissu mammaire au cours de la gestation, l'autre en comparaison deux à deux confrontant les gènes différemment exprimés entre le début de la gestation avec la lactation et entre les deux génotypes au locus *DGAT1*. L'analyse des données a été effectuée grâce à des collaborations étroites avec des biostatisticiens de l'UMR 518 d'AgroParisTech et de la Station d'Amélioration Génétique des Animaux (SAGA) à l'Inra de Toulouse, permettant la conception de différentes méthodes d'analyse des données. Ainsi, la fonction « *normalisation* » du package « anapuce » développée sous R par J. Aubert a été adaptée aux fichiers de quantification du logiciel « Feature Extraction » (Agilent Technologies). La fonction « *verifdonnees* » a été crée afin de vérifier un fichier de données normalisées par la fonction « *normalisation* ». Les données normalisées et vérifiées ont été analysées soit par régression linéaire pour l'analyse cinétique, soit par analyse différentielle en utilisant la fonction « *DiffAnalysis* » du package « anapuce » pour les comparaisons deux à deux. Par ailleurs, nous avons ordonné les gènes selon l'intensité de leur signal, pour évaluer leur niveau d'expression dans différents états physiologiques du tissu mammaire. De plus, les gènes dont l'expression variait entre plusieurs conditions ont été soit regroupés en profil de même expression dans le cas de l'étude cinétique, soit analysés en ACP, afin de déterminer ceux qui discriminaient le plus les deux génotypes au locus *DGAT1*.

Pour l'interprétation biologique des données, nous avons d'abord utilisé le logiciel « DAVID » (Dennis *et al.*, 2003), accessible gratuitement sur Internet ; puis le logiciel « Ingenuity Pathway Analysis », accessible sous licence, mais proposant une plus grande fonctionnalité dans l'analyse des données.

Mise à part les normes MIAME (« Minimum Information About a Microarray Experiment ») pour la publication des données de puces à ADN (Brazma *et al.*, 2001), il n'y a pas de règles établies pour l'analyse des données. Ainsi, l'analyse d'un ensemble de données par différents laboratoires et méthodes n'aboutira pas forcément aux mêmes conclusions en terme de nombre de gènes différemment exprimés. Dans ce travail de thèse, pour chaque projet, différentes méthodes d'analyse ont été expérimentées. A chaque fois, les analyses procuraient une hiérarchie comparable des gènes différemment exprimés. Il a donc été décidé, pour la présentation du matériel et des résultats, de n'utiliser que des outils accessibles par la communauté scientifique (des fonctions publiées sur R par exemple).

Réseaux de gènes modifiés lors de la différenciation terminale de la glande mammaire caprine

L'étude de la différenciation terminale de la glande mammaire caprine au moyen d'une analyse transcriptomique comparant cinq stades de développement (4 stades de gestation et un stade de lactation) a été effectuée en deux étapes. La première a consisté à comparer directement deux stades supposés biologiquement très différents : le début de la gestation (46 jours de gestation) et la lactation, à 40 jours, ce qui correspond à un stade de lactation bien établi (comparaison G46-L40). La deuxième étape a comparé cinq stades de développement en une étude cinétique.

De la première analyse, il est ressorti une liste de gènes surexprimés dans la glande mammaire en lactation et en début de gestation. La liste sur la glande mammaire lactante contient tous les gènes déjà identifiés comme intervenant dans la synthèse des constituants du lait, à savoir les gènes spécifiant les lactoprotéines (les caséines, l'α-lactalbumine et la β-lactoglobuline) et les gènes spécifiant les enzymes et les protéines de la synthèse et de la sécrétion des lipides du lait (leptine, stéaroyl CoA désaturase, lipoprotéine lipase, « fatty acid binding protein », glycérol-3-phosphate acyltransférase, butyrophiline, mucine 15 et « milk fat globule-EGF factor 8 »). La liste sur la glande mammaire caprine en début de gestation (G46) contient des gènes représentatifs du tissu conjonctif (collagène et thymosine), des cellules myoépithéliales (actine, myosine et kératine) et de la CEM (caséine β), ce qui indique que même en tout début de gestation, les tissus de la glande sont déjà très hétérogènes. Dans cette étude, l'hypothèse que les deux glandes comparées sont extrêmement différentes a également été validée, confirmant ainsi le plan expérimental de la deuxième étape de ce travail, où la comparaison des cinq stades de développement n'effectuait pas la comparaison directe du début de la gestation (G46) avec la lactation (L40).

La deuxième analyse de ce projet a tracé les profils d'expression des gènes sur cinq stades de développement, avec pour objectif de mieux comprendre les mécanismes et les facteurs locaux participant à la différenciation terminale de la glande mammaire caprine. Cette analyse a permis l'identification de 1 696 gènes dont l'expression varie significativement entre les cinq stades comparés. Les gènes dont l'expression diminue au cours de la gestation et en lactation spécifient des protéines, telles que l'adiponectine et PPARγ, caractérisant les adipocytes. La surexpression à 70 et 110 jours de gestation des gènes de la réponse immunitaire innée et spécifique, tend à suggérer que les tissus mammaires se caractériseraient par un tissu immunitaire sécrétant des facteurs d'attraction des cellules immunitaires (macrophages) qui participent au remodelage des tissus, à savoir la destruction

des adipocytes et la mise en place et le développement des cellules épithéliales. La différenciation ultime du tissu mammaire en tissu sécrétoire n'apparaît qu'après la parturition, et se caractérise, entre autres, par la surexpression des gènes spécifiant les acteurs de la synthèse et de la sécrétion des lipides du lait. L'intervention des gènes spécifiant des protéines de l'immunité innée dans le remodelage des tissus mammaires caprins pendant la gestation renforce l'hypothèse de Vorbach *et al.*, (2006) qui considère que la glande mammaire aurait pour origine fonctionnelle le système immunitaire inné et que la fonction première du lait serait une fonction protectrice avant de remplir une fonction nutritionnelle. La lactation pourrait, selon ces auteurs, trouver son origine dans un processus qui partage un certain nombre de voies de signalisation avec la réaction inflammatoire. La première partie de la gestation apparaît cruciale dans le processus de différenciation terminale de la glande mammaire des ruminants, avec l'idée que le déclenchement de la différenciation aurait lieu en tout début de gestation. Les gènes de l'immunité participeraient activement à ce processus de différenciation.

Effet du polymorphisme au locus DGAT1 sur l'activité des tissus mammaires bovins et les caractéristiques du lait

L'étude sur l'effet du polymorphisme K232A au locus *DGAT1* sur les performances laitières, la composition et les caractéristiques des structures supramoléculaires du lait (globules gras et micelles de caséines) et l'activité transcriptionnelle de la glande mammaire bovine, a révélé un impact profond de cette mutation sur le métabolisme des tissus mammaires.

Ces résultats ont été obtenus à partir d'un dispositif expérimental bien défini et contrôlé, dans lequel tous les animaux étaient homozygotes pour la mutation K232A, étaient élevés dans une même installation expérimentale, recevaient le même régime alimentaire et étaient au même stade de lactation au moment des prélèvements. A chaque animal de génotype GC/GC (A232), correspondait soit sa pleine sœur, soit sa demi-sœur de génotype AA/AA (K232), constituant *in fine* 6 couples de vaches. Les animaux étaient issus de la deuxième génération d'un croisement Prim'Holstein x Normand à l'origine du dispositif de détection de QTL laitiers du Pin-au-Haras (Larroque *et al.*, 2003).

L'analyse des performances laitières et de la composition du lait a confirmé l'effet de la mutation K232A sur le taux butyreux (TB) et la composition en acides gras (AG) des laits. A savoir une baisse significative du TB et des AG à chaîne moyenne et insaturés (C14:1 et C16:1) et une augmentation des AG à chaîne moyenne et saturée et longue et insaturée

(C14:0, C18:1 *cis11*, C18:1 *cis12, C18:2 n-6 et* C18:3 *n-3*) avec le variant A232. Les globules gras étaient également plus petits avec ce variant.

Le faible effectif d'individus, n'a pas permis, dans cette expérience, de mettre en évidence une relation entre la mutation K232A et les teneurs en protéines majeures, puisqu'il était impossible de dissocier le génotype au locus des protéines majeures du génotype au locus *DGAT1* et que de nombreux facteurs affectent la quantité des protéines majeures.

L'analyse du transcriptome à partir d'ARN extraits des biopsies de glande mammaire a montré que les gènes spécifiant les acteurs des voies de synthèse du lactose, des lipides et des protéines étaient affectées par le polymorphisme K232A dans le sens d'une surexpression chez les individus de génotype GC/GC (A232). Ces observations semblent indiquer que les tissus mammaires de ces individus présenteraient une sensibilité accrue à une hormone métabolique : l'insuline. Cette augmentation de sensibilité a déjà été observée dans le tissu adipeux de souris invalidées (KO) pour le gène *DGAT1* (Chen *et al.*, 2002). Ces résultats et cette hypothèse restent cependant à confirmer sur un effectif d'animaux plus important et à valider en analysant de façon plus précise l'expression de gènes clés du métabolisme des glucides et des acides gras, mais aussi des protéines (en favorisant la captation des acides aminés).

Confirmation de l'importance de certains gènes dans la différenciation et le fonctionnement de la glande mammaire

Les études conduites sur la différenciation terminale de la glande mammaire caprine et sur l'effet du polymorphisme au locus *DGAT1* sur le fonctionnement de la glande mammaire bovine ont permis de répertorier quelques gènes dont l'expression est naturellement affectée par des changements de l'état physiologique de la glande mammaire.

L'enzyme DGAT1 est nécessaire au développement normal de la glande mammaire, puisque son invalidation entraîne une diminution de la prolifération des cellules épithéliales dans les tissus mammaires murins (Cases *et al.*, 2004). Notre étude sur la différenciation terminale des tissus mammaires caprins, a révélé que le gène DGAT1 est exprimé tout au long de la gestation avec une forte surexpression en lactation. De plus, la mutation K232A au locus *DGAT1* affecte toutes les voies de synthèse des constituants du lait dans les tissus mammaires bovins, soulignant le rôle essentiel de l'enzyme codée par ce gène dans le fonctionnement de la cellule épithéliale, la production et la sécrétion du lait.

Notre étude sur les chèvres confirme également le rôle de la caséine α_{s1} dans les mécanismes de sécrétion des constituants du lait. En effet, l'expression du gène *CSN1S1*,

spécifiant la caséine α_{s1} n'a été observée qu'en lactation, contrairement aux autres gènes spécifiant les lactoprotéines majeures (notamment les autres caséines) dont certains sont exprimés de façon significative à mi-gestation. Cette observation est en cohérence avec l'idée que les mécanismes de sécrétion des constituants du lait seraient activés uniquement à la parturition et dépendraient, pour ce qui concerne les micelles de caséines, de la présence de la caséine α_{s1}. (Chanat *et al.*, 1999) avaient déjà souligné l'importance de cette caséine dans les mécanismes de sécrétion des caséines. Par ailleurs, dans cette étude, la biosynthèse et la sécrétion des lipides du lait ne sont également activées qu'à la parturition. Ce travail, confirme donc le lien entre synthèse et sécrétion des lipides et des protéines du lait, lien déjà établi par l'étude du polymorphisme au locus caséine α_{s1} chez la chèvre (Grosclaude *et al.*, 1994).

L'étude sur la mutation K232A décrit un doublement du nombre de transcrits du gène spécifiant l'α-lactalbumine avec le variant GC (A232), en lien avec une tendance à l'augmentation du taux de lactose et de la quantité de lactose sécrétée. Cette observation renforce l'idée selon laquelle cette protéine qui modifie l'activité de l'UDP-Galactosyl transférase pour favoriser son activité lactose-synthase est un élément régulateur déterminant de la quantité de lactose sécrétée dans le lait (Holsinger, 1997).

Le gène spécifiant le transporteur moléculaire ABCG2 est faiblement exprimé à mi-gestation et fortement en lactation dans les tissus mammaires caprins (données non publiées). De plus, c'est un des gènes qui discrimine le plus le génotype au locus *DGAT1* chez la vache. Or, une mutation au locus *ABCG2* a été associée à une modification des taux protéiques et butyreux et de la quantité de lait (Cohen-Zinder *et al.*, 2005; Olsen *et al.*, 2007). Au delà de son rôle de transporteur et de protecteur contre les xénobiotiques (Robey *et al.*, 2009), ABCG2 apparaît, dans notre étude, comme un facteur affectant les synthèses protéiques et lipidiques dans la glande mammaire.

Perspectives

Le travail réalisé au cours de cette thèse n'est qu'une amorce dans la compréhension des mécanismes de différenciation et de fonctionnement de la glande mammaire. Il a été souligné l'implication de facteurs de l'immunité innée dans la différenciation des tissus mammaires caprins, et renforcé la notion de l'existence d'un lien étroit entre les voies de synthèse des lipides, protéines et lactose dans les tissus mammaires bovins.

Toutefois, l'approche utilisée ici est une approche essentiellement transcriptomique qui étudie uniquement les transcrits dans un tissu, c'est-à-dire les ARN, caractérisant

l'expression du gène mais pas la fonctionnalité et la quantité des protéines traduites. Il est donc nécessaire de compléter cette approche par d'autres méthodes d'analyse des tissus, telles que la protéomique, la métabolomique ou la biologie cellulaire pour confirmer les observations.

Par ailleurs, l'annotation des sondes présentent sur les puces à ADN est encore très partielle. Une avancée dans cette voie permettrait d'affiner l'analyse, de préciser et de compléter certains réseaux de gènes.

De la confrontation des résultats issus des deux études auxquelles j'ai participé, il ressort quelques gènes clés dont l'implication dans la différenciation et le fonctionnement de la glande mammaire reste à approfondir. Il serait intéressant de comparer les répertoires de gènes établis dans ce travail avec d'autres répertoires de gènes issus d'analyses transcriptomiques réalisées sur la glande mammaire, mais dans d'autres espèces. Ceci pourrait être fait par l'intermédiaire de la base de données mise en place par l'International Milk Genomics Consortium, qui s'est notamment donné cet objectif.

A la faveur de l'étude portant sur les effets de la mutation K232A de l'enzyme DGAT1, nous avons tenté de mettre en relation des données de transcriptomique avec des données phénotypiques (performances laitières, composition et caractéristiques des structures supramoléculaires du lait). Cet essai n'a pas été totalement probant, puisque qu'il utilisait une technique « maison », à partir d'une liste de gènes sélectionnés (gènes différemment exprimés). Il est envisageable de renouveler l'expérience, en effectuant une analyse discriminante utilisant des outils, non disponibles lors de notre étude, dédiés à la mise en relation de données phénotypiques avec des données de transcriptomique tel que le « sparse PLS regression » (Lê Cao *et al.*, 2008).

L'unité productrice du lait est la cellule épithéliale mammaire. Or, cette cellule fonctionne en lien étroit avec son environnement, le stroma, structure fibreuse conjonctive qui contient des fibroblastes et d'autres types cellulaires. Les analyses transcriptomiques présentées ici ont toutes porté sur le tissu mammaire, pris dans sa globalité. Il serait intéressant d'analyser plus précisément à quels types cellulaires se rapportent les gènes dont l'expression est modulée dans les différents états physiologiques étudiés. Ceci pourrait être réalisé en utilisant la technologie de microdissection laser qui permet de sélectionner, pour isoler et analyser, un type cellulaire à partir d'une coupe de tissu (Emmert-Buck *et al.*, 1996). Une telle démarche a déjà été engagée dans l'équipe où ce travail de thèse a été réalisé.

Bibliographie

A

Akers, R.M. (1985) Lactogenic hormones: binding sites, mammary growth, secretory cell differentiation, and milk biosynthesis in ruminants. Journal of Dairy Science 68(2), 501-519.

Akers, R.M. (1990) Lactational physiology: a ruminant animal perspective. Protoplasma 159, 96-111.

Akers, R.M., McFadden, T.B., Purup, S., Vestergaard, M., Sejrsen, K. & Capuco, A.V. (2000) Local IGF-I axis in peripubertal ruminant mammary development. Journal of Mammary Gland Biology and Neoplasia 5(1), 43-51.

Anderson, R.R., Harness, J.R., Snead, A.F. & Salah, M.S. (1981) Mammary growth pattern in goats during pregnancy and lactation. Journal of Dairy Science 64, 427-432.

Anderson, S.M., Rudolph, M.C., McManaman, J.L. & Neville, M.C. (2007) Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Research 9(1), 204:218.

Andersson, M., Wettesten, M., Boren, J., Magnusson, A., Sjöberg, A., Rustaeus, S. & Olofsson, S.O. (1994) Purification of diacylglycerol:acyltransferase from rat liver to near homogeneity. Journal of Lipid Research 35, 535-545.

B

Baldwin, R.L. (1966) Enzymatic activities in mammary glands of several species. Journal of Dairy Science 49, 1533-1542.

Ballester, M., Sanchez, A. & Folch, J.M. (2005) Polymorphisms in the goat beta-lactoglobulin gene. Journal of Dairy Research 72, 379-384.

Bar-Or, C., Czosnek, H. & Koltai, H. (2007) Cross-species microarray hybridizations: a developing tool for studying species diversity. TRENDS in Genetics 23(4), 200-207.

Bauman, D.E., Mather, I.H., Wall, R.J. & Lock, A.L. (2006) Major advances associated with the biosynthesis of milk. Journal of Dairy Science 89, 1235-1243.

Bell, R.M., Ballas, L.M. & Coleman, R.A. (1981) Lipid topogenesis. Journal of Lipid Research 22, 391-403.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Serie B 47(1), 289-300.

Bennewitz, J., Reinsch, N., Paul, S., Looft, C., Kaupe, B., Weimann, C., Erhardt, G., Thaller, G., Kühn, C., Schwerin, M., Thomsen, H., Reinhardt, F., Reents, R. & Kalm, E. (2004) The DGAT1 K232A Mutation Is Not Solely Responsible for the Milk Production Quantitative Trait Locus on the Bovine Chromosome 14. Journal of Dairy Science 87, 431-442.

Bernard, L., Leroux, C. & Chilliard, Y. (2008) Expression and nutritional regulation of lipogenic genes in the ruminant lactating mammary gland. Advances in Experimental Medicine and Biology 606, 67-108.

Bevilacqua, C., Helbling, J.C., Miranda, G. & Martin, P. (2006) Translational efficiency of casein transcripts in the mammary tissue of lactating ruminants. Reproduction Nutrition Development 5, 567-578.

Bidanel, J.P., Boichard, D. & Chevalet, C. (2008) De la génétique à la génomique. INRA Productions Animales 21(1), 15-32.

Bionaz, M. & Loor, J.J. (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9, 366.

Bogard, M., Ameziane, N. & Lamoril, J. (2008) Microarray d'ADN et profils d'expression des gènes Première partie: concept, fabrication et mise en oeuvre Immuno-analyse et biologie spécialisée 23, 71-88.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M.Y. & Levéziel, H. (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics Selection Evolution 35, 77-101.

Bole-Fyesot, C., Goffin, V., Edery, M., Binart, N. & Kelly, P. (1998) Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. Endocrine Reviews 19(3), 225-268.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. & Vingron, M. (2001) Minimum information about microarray experiment (MIAME) - toward standards for microarray data. Nature Genomics 29, 365-371.

Breckenridge, W.C. & Kuksis, A. (1968) Specific distribution of short-chain fatty acids in molecular distillates of bovine milk fat. Journal of Lipid Research 9, 388-393.

Briard, V., Leconte, N., Michel, F. & Michalski, M.C. (2003) The fatty acid composition of small and large naturally occuring milk fat globules. European Journal of Lipid Science and Technology 105, 677-682.

Brown, D.A. (2001) Lipid droplets: proteins floating on a pool of fat. Current Biology 11, R446-R449.

Butte, A. (2002) The use and analysis of microarray data. Nature Review 1, 951-960.

C

Capuco, A.V., Connor, E.E. & Wood, D.L. (2008) Regulation of mammary gland sensitivity to thyroid hormones during the transition from pregnancy to lactation. Experimental Biology and Medecine 233(10), 1309-1314.

Cario, E. & Podolsky, D.K. (2000) Differential alteration in intestinal epithelial cell expression of toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. Infection and Immunity 68(12), 7010-7017.

Caroli, A., Chiatti, F., Chessa, S., Rignanese, D., Ibeagha-Awemu, E.M. & Erhardt, G. (2007) Characterization of the casein gene complex in West African goats and description of a new alpha S1 casein polymorphism. Journal of Dairy Science 90(6), 2989-2996.

Cases, S., Smith, S.J., Zheng, Y.W., Myers, H.M., Lear, S.R., Sande, E., Novak, S., Collins, C., Welch, C.B., Lusis, A.J., Erickson, S.K. & Farese, R.V. (1998) Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proceedings of the National Academy of Sciences of the United States of America 95, 13018-13023.

Cases, S., Stones, S.J., Zhou, P., Yen, E., Tow, B., Lardizabal, K.D., Voelker, T. & Farese, R.V. (2001) Cloning of DGAT2, a Second Mammalian Diacylglycerol Acyltransferase, and Related Family Members. The Journal of Biological Chemistry 276(42), 38870-38876.

Cases, S., Zhou, P., Shillingford, J.M., Wiseman, B.S., Fish, J.D., Angle, C.S., Hennighausen, L., Werb, Z. & Farese, R.V. (2004) Development of the mammary gland requires DGAT1 expression in stromal and epithelial tissues. Develoment 131, 3047-3055.

Casper, D.P., Schingoethe, D.J., Middaugh, R.P. & Baer, R.J. (1988) Lactational responses of dairy cows to diets containing regular and high oleic acid sunflower seeds. Journal of Dairy Science 71(5), 1267-1274.

Ceriotti, G., Chessa, S., Bolla, P., Budelli, E., Bianchi, L., Duranti, E. & Caroli, A. (2004) Single nucleotide polymorphisms in the Ovine casein genes detected by polymerase chain reaction-single strand conformation polymorphism. Journal of Dairy Science 87(8), 2606-2613.

Chanat, E., Martin, P. & Ollivier-Bousquet, M. (1999) AlphaS1-casein is required for the efficient transport of beta-casein and kappa-casein from the endoplasmic reticulum to the Golgi apparatus mammary epithelial cells. Journal of Cell Science 112, 3399-3412.

Chen, H.C. (2006) Enhancing energy and glucose metabolism by disrupting triglyceride synthesis: lessons from mice lacking DGAT1. Nutrition and Metabolism 3(10).

Chen, H.C., Jensen, D.R., Myers, H.M., Eckel, R.H. & Farese, R.V.J. (2003) Obesity resistance and enhanced glucose metabolism in mice transplanted with white adipose tissue lacking acyl-CoA:diacylglycerol acyltransferase 1. The Journal of Clinical Investigation 111(11), 1715-1722.

Chen, H.C., Smith, S.J., Ladha, Z., Jensen, D.R., Ferreira, L.D., Pulawa, L.K., McGuire, J.G., Pitas, R.E., Eckel, R.H. & Farese, R.V. (2002) Increased insulin and leptin sensitivity in mice lacking acylCoA:diacylglycerol acyltransferase 1. The Journal of Clinical Investigation 109, 1049-1055.

Cheng, D., Meegalla, R.L., Bokand, H.E., Cromley, D.A., Billheimer, J.T. & Young, P.R. (2001) Human acyl-CoA:diacylglycerol acyltransferase is a tetrameric protein. Biochemical Journal 359, 707-714.

Chessa, S., Rignanese, D., Chiatti, F., Radeghieri, A., Gigliotti, C. & Caroli, A. (2008a) Technical note: simultaneous identification of CSN1S2 A, B, C, and E alleles in goats by polymerase chain reaction-single strand conformation polymorphism. Journal of Dairy Science 91(3), 1214-1217.

Chessa, S., Rignanese, D., Küpper, J., Pagnacco, G., Erhardt, G. & Caroli, A. (2008b) Short communication: the beta-casein (CSN2) silent allele C1 is highly spread in goat breeds. Journal of Dairy Science 91(11), 4433-4436.

Chianese, L., Garro, G., Mauriello, R., Laezza, P., Ferranti, P. & Addeo, F. (1996) Occurence of five alpha s1-casein variants in ovine milk. Journal of Dairy Research 63(1), 49-59.

Chiatti, F., Chessa, S., Bolla, P., Cigalino, G., Caroli, A. & Pagnacco, G. (2007) Effect of kappa-casein polymorphism on milk composition in the Orobica goat. Journal of Dairy Science 90(4), 1962-1966.

Chilliard, Y. & Ferlay, A. (2004) Dietary lipids and forages interactions on cow and goat milk fatty acid composition and sensory properties. Reproduction Nutrition Development 44, 467-492.

Chilliard, Y., Ferlay, A., Mansbridge, R.M. & Doreau, M. (2000) Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, trans and conjugated fatty acids. Annales de Zootechnie 49, 181-205.

Chilliard, Y., Rouel, J. & Leroux, C. (2006) Goat's alpha-s1 casein genotype influences its milk fatty acid composition and delta-9 desaturation ratios. Animal Feed Science and Technology 131, 474-487.

Choi, K.M., Barash, I. & Rhoads, R.E. (2004) Insulin and prolactin synergistically stimulate beta-casein messenger ribonucleic acid translation by cytoplasmic polyadenylation. Molecular Endocrinology 18(7), 1670-1686.

Chrétien, M. (1995) La chimie de la vie. In La biologie: Campbell, DeBoeck Université.

Clarkson, R.W.E., Wayland, M.T., Lee, J.H., Freeman, T. & Watson, C.J. (2003) Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Research 6(2), R92-R109.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.H., Drackley, J.K., Band, M.R., Hernandez, A.G., Shani, M., Lewin, H.A., Weller, J.I. & Ron, M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Research 15, 936-944.

Coleman, R. & Bell, R.M. (1976) Triacylglycerol Synthesis in Isolated Fat Cells - Studies on the microsomal diacylglycerol acyltransferase activity using ethanol-dispersed diacylglycerols. The Journal of Biological Chemistry 251(15), 4537-4543.

Coleman, R. & Bell, R.M. (1978) Evidence that biosynthesis of phosphatidylethanolamine, phosphatidylcholine, and triacylglycerol occurs on the cytoplasmic side of microsomal vesicles. The Journal of Cell Biology 76, 245-253.

Coleman, R.A., Lewin, T.M. & Muoio, D.M. (2000) Physiological and nutritional regulation of enzymes of triacylglycerol synthesis. Annual Review of Nutrition 20, 77-103.

Collomb, M., Sollberger, H., Bütikofer, U., Sieber, R., Stoll, W. & Schaeren, W. (2004) Impact of a basal diet of hay and fodder beet supplemented with rapeseed, linseed and sunflowerseed on the fatty acid composition of milk fat. International Dairy Journal 14, 549-559.

Connor, E.E., Meyer, M.J., Li, R.W., Van Amburgh, M.E., Boisclair, Y.R. & Capuco, A.V. (2007) Regulation of gene expression in the bovine mammary gland by ovarian steroids. Journal of Dairy Science 90(E. Suppl.), E55-E65.

Coppieters, W., Riquet, J., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D. & Georges, M. (1998) A QTL with major effect on milk yield and composition maps to bovine chromosome 14. Mammalian Genome 9, 540-544.

Cosenza, G., Gallo, D., Illario, R., di Gregorio, P., Senese, C., Ferrara, L. & Ramunno, L. (2003) A Mval PCR-RFLP detecting a silent allele at the goat alpha-lactalbumin locus. Journal of Dairy Research 70, 355-357.

Couvreur, S. & Hurtaud, C. (2007) Le globule gras du lait : sécrétion, composition, fonctions et facteurs de variation. INRA Productions Animales 20(5), 369-382.

Couvreur, S., Hurtaud, C., Delaby, L., Michel, F. & Peyraud, J.L. (2003) Effet d'un régime à base d'herbe mi-fanée ou d'ensilage de maïs associé ou non à un déficit énergétique sur les propriétés des globules gras du lait. Rencontres, Recherches, Ruminants 10, 235.

D

De Kruif, C.G. & Holt, C. (2003) Casein micelle structure, functions and interactions. In Advances Dairy Chemistry - Volume 1 - Proteins - Third edition Part A: Fox, P.F. and McSweeney, P.H.L.

Decean, C. & Adda, J. (1970) Evolution de la sécrétion des acides gras des matières grasses du lait au cours de la lactation de la vache. Ann. Biol. anim. Bioch. Biophys. 10(4), 659-677.

Delmar, P., Robin, S. & Daudin, J.J. (2005) VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data. Bioinformatics 21(4), 502-508.

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. & Lempicki, R.A. (2003) DAVID: database for anotation, visualization, and integreted discovery. Genome Biology 4(P3).

Doreau, M. & Ferlay, A. (1994) Digestion and utilisation of fatty acids by ruminants. Animal Feed Science and Technology 45, 379-396.

E

Ebner, K.E. (1971) Biosynthesis of lactose. Journal of Dairy Science 54(8), 1229-1233.

Ellis, S.E. (1998) Mechanism controlling ductal morphogenesis in the ruminant mammary gland.

Emadi, S., Kasturirangan, S., Wang, M.S., Schulz, P. & Sierks, M.R. (2009) Detecting morphologically distinct oligomeric forms of alpha-synuclein. The Journal of Biological Chemistry 284, 11048-11058.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A. & Liotta, L.A. (1996) Laser capture microdissection. Science 8(274(5289)), 998-1001.

Erb, R.E. (1977) Hormonal control of mammogenesis and onset of lactation in cows - A review. Journal of Dairy Science 60(2), 155-169.

Evers, J.M. (2004) The milkfat globule membrane - compositional and structural changes post secretion by the mammary secretory cell. International Dairy Journal 14, 661-674.

Everts, R.E., Chavatte-Palmer, P., Razzak, A., Hue, I., Green, C.A., Oliveira, R., Vignon, X., Rodriguez-Zas, S.L., Cindy Tian, X., Yang, X., Renard, J.P. & Lewin, H.A. (2008) Aberrant gene expression patterns in placentomes are associated with phenotypically normal and abnormal cattle cloned by somatic cell nuclear transfer. Physiological Genomics 33, 65-77.

F

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilkki, J. & Georges, M. (2002) Simultaneous Mining of Linkage and Linkage Disequilibrium to Fine Map Quantitative Trait Loci in Outbred Half-Sib Pedigrees: Revisiting the Location of a Quantitative Trait Locus with Major Effect on Milk Production on Bovine Chromosome 14. Genetics 161, 275-287.

Farr, V.C., Stelwagen, K., Cate, L.R., Molenaar, A.J., McFadden, T.B. & Davis, S.R. (1996) An improved method for the routine biopsy of bovine mammary tissue. Journal of Dairy Science 79, 543-549.

Farrell, H.M., Jimenez-Flores, R., Bleck, G.T., Brown, E.M., Butler, J.E., Creamer, L.K., Hicks, C.L., Hollar, C.M., Ng-Kwai-Hang, K.F. & Swaisgood, H.E. (2004) Nomenclature of the proteins of cow's milk - Sixth revision. Journal of Dairy Science 87(6), 1641-1674.

Finucane, K.A., McFadden, T.B., Pond, J.P., Kennelly, J.J. & Zhao, F.Q. (2008) Onset of lactation in the bovine mammary gland: gene expression profiling indicates a strong inhibition of gene expression in cell proliferation. Functional and Integrative Genomics.

Frandson, R.D. (1986) Anatomy and physiology of farm animals. 4th edition, Lea and Febiger, Philadelphia.

G

Gautier, M., Capitan, A., Fritz, S., Eggen, A., Boichard, D. & Druet, T. (2007) Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphism in french dairy cattle. Journal of Dairy Science 90, 2980-2988.

Gerson, T., Shorland, F.B., Wilson, G.F. & Reid, C.W.S. (1967) Origin of glyceride fatty acids in cow milk fat. Journal of Dairy Science 51(3), 356-361.

Gonthier, C., Mustafa, A.F., Ouellet, D.R., Chouinard, P.Y., Berthiaume, R. & Petit, H.V. (2005) Feeding micronized and extruded flaxseed to dairy cows: effects on blood parameters and milk fatty acid composition. Journal of Dairy Science 88(2), 748-756.

Goudon-Evans, V., Lin, E.Y. & Pollard, J.W. (2002) Requirement of macrophages and eosinophils and their cytokines/chemokines for mammary gland development. Breast Cancer Research 4, 155-164.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. (2002) Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. Genome Research 12, 222-231.

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J.J., Kvasz, A., Mni, M., Simon, P., Frère, J.M., Coppieters, W. & Georges, M. (2004) Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proceedings of the National Academy of Sciences of the United States of America 101(8), 2398-2403.

Grosclaude, F., Ricordeau, G., Martin, P., Remeuf, F., Vassal, L. & Bouillon, J. (1994) Du gène au fromage : le polymorphisme de la caséine alpha S1 caprine, ses effets, son evolution. INRA Productions Animales 7(1), 3-19.

Guinard-Flament, J., Delamaire, E., Lemosquet, S., Boutinaud, M. & David, Y. (2006) Changes in mammary uptake and metabolic fate of glucose with once-daily milking and feed restriction in dairy cows. Reproduction Nutrition Development 5, 589-598.

H

Haagsman, H.P., DeHaas, C.G.M., Geelen, M.J.H. & vanGolde, L.M.G. (1982) Regulation of Triacylglycerol Synthesis in the Liver - Modulation of diacylglycerol acyltransferase activity in vitro. The Journal of Biological Chemistry 257(18), 10593-10598.

Hawke, J.C. & Taylor, M.W. (1995) Influence of nutritional factors on the yield, composition and physical properties of milk fat. In Advanced Dairy Chemistry-2 : Lipids, pp. 37-88. Edited by P.F. Box.

Heid, H.W. & Keenan, T.W. (2005) Intracellular origin and secretion of milk fat globules. European Journal of Biochemistry 84, 245-258.

Hennighausen, L. & Robinson, G.W. (1998) Think globally, act locally: the making of a mouse mammary gland. Genes and Development 12, 449-455.

Hennighausen, L. & Robinson, G.W. (2001) Signaling Pathways in Mammary Gland Development. Developmental Cell 1, 467-475.

Heyen, D.W., Weller, J.I., Ron, M., Band, M., Beever, J.E., Feldmesser, E., Da, Y., Wiggans, G.R., Van Raden, P.M. & Lewin, H.A. (1999) A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiological Genomics 1, 165-175.

Holsinger, V.H. (1997) Physical and chemical properties of lactose. In Advanced dairy chemistry volume 3: Lactose, water, salts and vitamins. Edited by P.F. Fox. Chapman and Hall.

Holt, C. (2004) An equilibrium thermodynamic model of the sequestration of calcium phosphate by casein micelles and its application to the calculation of the partition of salts in milk. European Biophysics Journal 33, 421-434.

Hovey, R.C., Auldist, D.E., Mackenzie, D.D. & McFadden, T.B. (2000) Preparation of an epithelium-free mammary fat pad and subsequent mammogenesis in ewes. Journal of Animal Science 78, 2177-2185.

Hovey, R.C., McFadden, T.B. & Akers, R.M. (1999) Regulation of mammary gland growth and morphogenesis by the mammary fat pad: a species comparison. Journal of Mammary Gland Biology and Neoplasia 4(1), 53-68.

Hovey, R.C., Trott, J.F. & Vonderhaar, B.K. (2002) Establishing a framework for the functional mammary gland: from endicrinology to morphology. Journal of Mammary Gland Biology and Neoplasia 7, 17-38.

Hu, E., Liang, P. & Spiegelman, B.M. (1996) AdipoQ is a novel adipose-specific gene dysregulated in obesity. The Journal of Biological Chemistry 271(18), 10697-10703.

Hurtaud, C., Faucon, F. & Peyraud, J.L. (2006) Effet de différentes formes d'apport de lin et de colza dans l'alimentation des vaches laitières sur les propriétés physiques et sensorielles du beurre. In 13ème Rencontres Recherches Ruminants Paris.

Hurtaud, C., Goudedranche, H., Delaby, L., Camier-Caudron, B. & Peyraud, J.L. (2002) Effet de la nature du régime hivernal sur la qualité du beurre et de l'emmental. Rencontres, Recherches, Ruminants 9, 369.

J

Jammes, H. & Djiane, J. (1988) Le développement de la glande mammaire et son contrôle hormonal dans l'espèce bovine. INRA Production Animale 1(5), 299-310.

Jenness, R. (1986) Symposium: species variation in mammary gland function. Lactational performance of various mammalian species. Journal of Dairy Science 69, 869-885.

Jiang, Z., De, S., Garcia, M.D., Griffin, K.B., Wu, X.L., Xiao, Q., Michal, J.J., Sharma, B.S. & Jansen, G.B. (2005) An independant confirmation of a quantitative trait locus for milk yield and composition traits on bovine chromosome 26. Journal of Animal Breeding and Genetics 122(4), 281-284.

Johnson, I.D. & Hart, I.C. (1986) Manipulation of milk yield with growth hormone. In Recent Advance in Animal Nutrition, pp. 105-123. Edited by L. Butterworths.

Jones, E.L., Shingfield, K.J., Kohen, C., Jones, A.K., Lupoli, B., Grandison, A.S., Beever, D.E., Williams, C.M., Calder, P.C. & Yaqoob, P. (2005) Chemical, physical and sensory properties of dairy products enriched with conjugated linoleic acid. Journal of Dairy Science 88, 2923-2937.

K

Kanazawa, T. & Kohmoto, K. (2002) Immunochemical demonstration of alphaS1 and beta casein in mouse mammary glands at ealy stages of pregnancy. The Journal of Histochemistry and Cytochemistry 50(2), 257-264.

Keenan, T.W. & Dylewski, P.D. (1995) Intracellular origin of milk lipid globules and the nature and structure of the milk lipid globule membrane. In Advanced Dairy Chemistry Volume 2: LipidsEdited by P.F. Fox. Chapman and Hall.

Kelsey, J.A., Corl, B.A., Collier, R.J. & Bauman, D.E. (2003) The effect of breed, parity, and stage of lactation on conjugated linoleic acid (CLA) in milk fat from dairy cows. Journal of Dairy Science 86(8), 2588-2597.

Kerr, M.K. (2003) Design Considerations for Efficient and Effective Microarray Studies. Biometrics 59, 822-828.

Kerr, M.K. & Churchill, G.A. (2001) Experimental design for gene expression microarrays. Biostatistics 2(2), 183-201.

Khaled, W.T., Read, E.K.C., Nicholson, S.E., Baxter, F.O., Brennan, A.J., Came, P.J., Sprigg, N., McKenzie, A.N.J. & Watson, C.J. (2007) The IL-4/IL-13/Stat6 signalling pathway promotes luminal mammary epithelial cell development. Development 134, 2739-2750.

Khatkar, M.S., Thomson, P.C., Tammen, I. & Raadsma, H.W. (2004) Quantitative trait loci mapping in the dairy cattle: review and meta-analysis. Genetics Selection Evolution 36, 163-190.

Kühn, C., Thaller, G., Winter, A., Bininda-Edmonds, O.R.P., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M. & Fries, R. (2004) Evidence for Multiple Alleles at the DGAT1 Locus Better Explains a Quantitative Trait Locus With Major Effect on Milk Far Content in Cattle. Genetics 167, 1873-1881.

Kuhn, N.J., Carrick, D.T. & Wilde, C.J. (1980) Symposium: milk synthesis - Lactose synthesis: the possibilities of regulation. Journal of Dairy Science 63(2), 328-336.

L

Labussière, J. (1993) Physiologie de l'éjection du lait. Conséquence sur la traite. In Biologie de la lactation: Martinet, J. et Houdebine, L.M., Les editions INSERM - INRA editions.

Lardizabal, K.D., Mai, J.T., Wagner, N.W., Wyrick, A., Voelker, T. & Hawkins, D.J. (2001) DGAT2 Is a New Diacylglycerol Acyltransferase Gene Family. The Journal of Biological Chemistry 276(42), 38862-38869.

Larroque, H., Gallard, Y., Boichard, D. & Colleau, J.J. (2003) Une expérimentation en croisement pour détecter des QTL chez les bovins laitiers. Rencontre Recherche Ruminants 10, 61.

Larson, B.L. (1985) Lactation. Iowa State University Press, Ames.

Lê Cao, K.A., Rossouw, D., Robert-Granié, C. & Besse, P. (2008) A sparse PLS for variable selection when integrating Omics data. Statistical Applications in Genetics and Molecular Biology 7(1), 1-32.

Lê, S., Josse, J. & Husson, F. (2008) FactoMineR: an R package for multivariate analysis. Journal of Statistical Software 25(1).

Lehner, R. & Kuksis, A. (1993) Triacylglycerol Synthesis by an sn-1,2(2,3)-Diacylglycerol Transacylase from Rat Intestinal Microsomes. The Journal of Biological Chemistry 268(12), 8781-8786.

Lehner, R. & Kuksis, A. (1996) Biosynthesis of Triacylglycerol. Progress in Lipid Research 35(2), 169-201.

Leung, Y.F. & Cavalieri, D. (2003) Fundamentals of cDNA microarray data analysis. TRENDS in Genetics 19(11), 649-659.

Leveziel, H. & Cribiu, E.P. (2000) 3- Cartographie des génomes. Etats des lieux de la cartographie du génome. INRA Productions Animales numéro hors Série, 103-107.

Lollivier, V., Guinard-Flament, J., Ollivier-Bousquet, M. & Marnet, P.G. (2002) Oxytocin and milk removal: two important sources of variation in milk production and milk quality during and between milkings. Reproduction Nutrition Developement 42(2), 173-186.

Long, J.C. & Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochemical Journal 417(1), 15-27.

Looft, C., Reinsch, N., Karall-Albrecht, C., Paul, S., Brink, M., Thomsen, H., Brockmann, G., Kühn, C., Schwerin, M. & Kalm, E. (2001) A mammary gland EST showing linkage desiquilibrium to a milk production QTL on bovine chromosome 14. Mammalian Genome 12, 646-650.

Lozeman, F.J., Middleton, C.K., Deng, J., Kazala, E.C., Verhaege, C., Mir, P.S., Laroche, A., Bailey, D.R.C. & Weselake, R.J. (2001) Characterization of microsomal diacylglycerol acyltransferase activity from bovine adipose and muscle tissue. Comparative Biochemistry and Physiology Part B 130, 105-115.

M

Man, W.C., Miyazaki, M., Chu, K. & Ntambi, J. (2006) Colocalization of SCD1 and DGAT2: implying preference for endogenous monounsaturated fatty acids in triglyceride synthesis. Journal of Lipid Research 47, 1928-1939.

Mann, J. & Skeaff, M. (2002) Chapter 3: Lipids. In Essentials of human nutrition - Second edition: Mann, J. and Truswell, Oxford University Press.

Marchin, S., Putaux, J.L., Pignon, F. & Léonil, J. (2007) Effects of the environmental factors on the casein micelle structure studied by cryo transmission electron microscopy and small-angle x-ray scattering/ultrasmall-angle x-ray scattering. The Journal of Chemical Physics 126(4), 045101.

Marshall, M.O. & Knudsen, J. (1976) Biosynthesis of Triacylglycerols Containning Short-Chain Fatty Acids in Lactating Cow Mammary Gland. European Journal of Biochemistry 81, 259-266.

Martin, P., Ferranti, P., Leroux, C. & Addeo, F. (2003) Non-bovine caseins: a quantitative variability and molecular diversity. In Advanced Dairy Chemistry - Volume 1 - Proteins - Third edition - Part A: Fox, P.F. and McSweeney, P.L.H.

Martin, P., Szymanowska, M., Zwierzchowski, L. & Leroux, C. (2002) The impact of genetic polymorphisms on the protein composition of ruminants milk. Reproduction Nutrition Development 42(5), 433-459.

Mary-Huard, T., Picard, F. & Robin, S. (2006) Introduction to statistical methods for microarray data analysis. In Mathematicals and computational methods in biology, pp. 155-230. Edited by Hermann. Paris.

Master, S.R., Hartman, J.L., D'Cruz, C.M., Moody, S.E., Keiper, E.A., Ha, S.I., Cox, J.D., Belka, G.K. & Chodosh, L.A. (2002) Functional microarray analysis of mammary organogenesis reveals a developmental role in adaptative thermogenesis. Molecular Endocrinology 16(6), 1185-1203.

Mather, I.H. (2000) A review and proposed nomenclature for major proteins of the milk-fat globule membrane. Journal of Dairy Science 83(2), 203-247.

Mather, I.H. & Keenan, T.W. (1998a) The cell biology of milk secretion: historical notes. Journal of Mammary Gland Biology and Neoplasia 3(3), 227-232.

Mather, I.H. & Keenan, T.W. (1998b) Origin and Secretion of Milk Lipids. Journal of Mammary Gland Biology and Neoplasia 3(3), 259-273.

Mayorek, N. & Bar-Tana, J. (1983) Medium Chain Fatty Acids as Specific Substrates dor Diglyceride Acyltransferase in Cultured Hepatocytes. The Journal of Biological Chemistry 258(11), 6789-6792.

Mayorek, N., Grinstein, I. & Bar-Tana, J. (1989) Triacylglycerol synthesis in cultured rat hepatocytes - The rate-limiting role of diacylglycerol acyltransferase. European Journal of Biochemistry 182, 395-400.

Meegalla, R.L., Billheimer, J.T. & Cheng, D. (2002) Concerted elevation of acyl-coenzyme A:diacylglycerol acyltransferase (DGAT) activity through independant stimulation of mRNA expression of DGAT1 and DGAT2 by carbohydrate and insulin. Biochemical and Biophysical Research Communications 298, 317-323.

Mellenberger, R.W. & Bauman, D.E. (1974) Metabolic adaptations during lactogenesis. Fatty acid synthesis in rabbit mammary tissue during pregnancy and lactation. Biochemical Journal 138, 373-379.

Mellenberger, R.W., Bauman, D.E. & Nelson, D.R. (1973) Metabolic adaptations during lactogenesis. Fatty acid and lactose synthesis in cow mammary tissue. Biochemical Journal 136, 741-748.

Mepham, T.B., Gaye, P., Martin, P. & Mercier, J.C. (1992) Biosynthesis of milk protein. In Advanced in dairy chemistry - Volume 1 - Proteins: Fox, P.F.

Mepham, T.B., Gaye, P. & Mercier, J.C. (1982) Biosynthesis of milk proteins. In Developments in Dairy Chemistry-1: Fox, P.F.

Michalski, M.C., Briard, V. & Michel, F. (2001) Optical parameters of milk fat globules for laser light scattering measurements. Le Lait 81, 787-796.

Mikkelsen, J. & Knudsen, J. (1987) Acyl-CoA-binding protein from cow - Binding characteristics and cellular and tissue distribution. Biochemical Journal 248, 709-714.

Miranda, G., Mahé, M.F., Leroux, C. & Martin, P. (2004) Proteomic tools to characterize the protein fraction of Equidae milk. Proteomics 4, 2496-2509.

Molenaar, A.J., Kuys, Y.M., Davis, S.R., Wilkins, R.J., Mead, P.E. & Tweedie, J.W. (1996) Elevation of lactoferrin gene expression in developing, ductal, resting, and regressing parenchymal epithelium of the ruminant mammary gland. Journal of Dairy Science 79, 1198-1208.

Mori, S., Nishikawa, S.I. & Yokota, Y. (2000) Lactation defect in mice lacking the helix-loop-helix inhibitor Id2. The EMBO Journal 19(21), 5772-5781.

Mulder, H. & Walstra, P. (1974) The milk fat globule. Center for Agricultural Publishing and Documentation, Wageningen, The Netherlands

N

Nadon, R. & Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis. TRENDS in Genetics 18(5), 265-271.

Nakhasi, H.L. & Qasba, P.K. (1979) Quantitation of milk proteins and their mRNAs in rat mammary gland at various stages of gestation and lactation. The Journal of Biological Chemistry 254(13), 69016-6025.

Naylor, M.J., Oakes, S.R., Gardiner-Garden, M., Harris, J., Blazek, K., Ho, T.W.C., Li, F.C., Wynick, D., Walker, A.M. & Ormandy, C.J. (2005) Trancriptional changes underlying the secretory activation phase of mammary gland development. Molecular Endocrinology 19(7), 1868-1883.

Nemeth, B.A., Tsang, S.W.Y., Geske, R.S. & Haney, P.M. (2000) Golgi targeting of the GLUT1 glucose transporter in lactating mouse mammary gland. Pediatric Research 47(4), 444-450.

Neveu, C. (2004) Effets du polymorphisme de la caseine alpha S1 sur la composition du lait et sur la structure de ses éléments figurés (micelles de caséines et globules gras) chez la chèvre. In Biologie Rennes: Université de Rennes 1.

Neveu, C., Riaublanc, A., Miranda, G., Chich, J.F. & Martin, P. (2002) Is the apocrine milk secretion process observed in the goat species rooted in the perturbation of the intracellular transport mechanism induced by defective alleles at the alpha S1 casein locus. Reproduction Nutrition Developement 42, 163-172.

Neville, M.C., McFadden, T.B. & Forsyth, I. (2002) Hormonal regulation of mammary differentiation and milk secretion. Journal of Mammary Gland Biology and Neoplasia 7(49-66).

Ng-Kwai-Hang, K.F. & Grosclaude, F. (2003) Genetic polymorphism of milk proteins. In Advanced Dairy Chemistry - Volume 1 - Proteins - Third edition -Part B, pp. 739-816. Edited by P.F.a.M. Fox, P.L.H.

Ng-Kwai-Hang, K.F., Hayes, J.F., Moxley, J.E. & Monardes, H.G. (1984) Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. Journal of Dairy Science 67(4), 835-840.

Nordstrom, E.A., Ryden, M., Backlund, E.C., Dahlman, I., Kaaman, M., Blomqvist, L., Cannon, B., Nedergaard, J. & Arner, P. (2005) A human-specific role of cell death-inducing DFFA (DNA fragmentation factor-alpha)-like effector A (CIDEA) in adipocyte lipolysis and obesity. Diabetes 54, 1726-1734.

O

O'Donnell, R., Holland, J.W., Deeth, H.C. & Alewood, P. (2004) Milk proteomics. International Dairy Journal 14, 1013-1023.

Ogg, S.L., Weldon, A.K., Dobbie, L., Smith, A.J.H. & Mather, I.H. (2004) Expression of butyrophilin (Btna1) in lactating mammary gland is essential for the regulated secretion of milk lipid droplets. Proceedings of the National Academy of Sciences 101(27), 10084-10089.

Ollier, S. (2007) Utilisation des puces à ADN pour l'étude de la régulation nutritionnelle du fonctionnement de la glande mammaire chez la chèvre en lactation. In Nutrition, p. 199. Clermont-Ferrand Université de Clermont-Ferrand I.

Ollier, S., Robert-Granié, C., Bernard, L., Chilliard, Y. & Leroux, C. (2007) Mammary transcriptome analysis of food-deprived lactating goats highlights genes involved in milk secretion and programmmed cell death. The Journal of Nutrition 137, 560-567.

Ollivier-Bousquet, M. (1993) Sécrétion des caséines : régulation hormonale. In Biologie de la lactation : Martinet, J. et Houdebine, L.M., Les éditions INSERM et INRA éditions.

Olsen, H.G., Nilsen, H., Hayes, B., Berg, P.R., Svendsen, M., Lien, S. & Meuwissen, T. (2007) Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. BMC Genetics 8(32).

Ozasa, S., Kempner, E.S. & Erickson, S.K. (1989) Functional size of acyl coenzyme A:diacylglycerol acyltransferase by radiation inactivation. Journal of Lipid Research 30, 1759-1762.

P

Palmiter, R.D. (1969) Hormonal induction and regulation of lactose synthetase in mouse mammary gland. Biochemical Journal 118, 409-417.

Palmquist, D.L. & Conrad, H.R. (1978) High fat rations for dairy cows. Effects of feed intake, milk and fat production, and plasma metabolites. Journal of Dairy Science 61(7), 890-901.

Pareek, C.S., Czarnik, U., Zabolewicz, T., Pareek, R.S. & Walawski, K. (2005) DGAT1 K232A quantitative trait nucleotide polymorphism in Polish Black-and-White cattle. Journal of Applied Genetics 46(1), 85-87.

Parodi, P.W. (1974) Variation in the fatty acid composition of milkfat: effet of stage of lactation. The Australian Journal of Dairy Technology, 145-148.

Pedchenko, V.K. & Imagawa, W. (2000) Pattern of expression of the KGF receptor and its ligands KGF and FGF-10 during postnatal mouse mammary gland development. Molecular reproduction and development 56, 441-447.

Peel, C.J. & Bauman, D.E. (1987) Somatotropin and lactation. Journal of Dairy Science 70(2), 474-486.

Peters, A. & Lamming, E. (1983) Hormone patterns and reproduction in cattle. In Practice - Farm Practice september 1983, 153-158.

Pitelka, D.R., Hamamoto, S.T., Duafala, J.G. & Nemanic, M.K. (1973) Cell contacts in the mouse mammary gland - Normal gland in postnatal development. The Journal of Cell Biology 56, 797-818.

Pointurier, H. & Adda, J. (1969) Beurrerie industrielle. La Maison Rustique, Paris.

Pougheon, S. & Goursaud, J. (2001) Le lait et ses constituants : caractéristiques physico-chimiques. In Lait, Nutrition, Santé, pp. 2-42. Edited by T.e. Doc.

Proudfoot, A.E.I., Buser, R., Borlat, F., Alouani, S., Soler, D., Offord, R.E., Schroder, J.M., Power, C.A. & Wells, T.N.C. (1999) Amino-terminallys modified RANTES

analogous demonstrate differential effects on RANTES receptors. The Journal of Biological Chemistry 274(45), 32478-32485.

Q

Quackenbush, J. (2001) Computational analysis of microarray data. Nature Review Genetics 2, 418-427.

Quinn, T. (1980) Dairy Farm Management. Delmar Publishers Alabama New York.

R

Rainard, P. & Poutrel, B. (1993) Protection immunitaire de la glande mammaire. In Biologie de la lactation: Martinet J. et Houdebine, L.M.

Rainard, P. & Riollet, C. (2006) Innate immunity of the bovine mammary gland. Veterinary Research 37, 369-400.

Rasmussen, J.T., Berglund, L., Rasmussen, M.S. & Petersen, T.E. (1998) Assignment of disulfide bridges in bovine CD36. European Journal of Biochemistry 257, 488-494.

Recio, I., Ramos, M. & Amigo, L. (1997) Study of the polymorphism of ovine alpha s1 and alpha s2 caseins by capillary electrophoresis. Journal of Dairy Research 64(4), 525-534.

Rego, O.A., Portugal, P.V., Sousa, M.B., Rosa, H.J.D., Vouzela, C.M., Borba, A.E.S. & Bessa, R.J.B. (2004) Effect of diet on the fatty acid pattern of milk from dairy cows. Animal Research 53, 213-220.

Riquet, J., Coppieters, W., Cambisano, N., Arranz, J.J., Berzi, P., Davis, S.K., Grisart, B., Farnir, F., Karim, L., Mni, M., Simon, P., Taylor, J.F., Vanmanshoven, P., Wagenaar, D., Womack, J.E. & Georges, M. (1999) Fine-mapping of quantitative trait loci by identity by descent in outbred populations: Application to milk production in dairy cattle. Proceedings of the National Academy of Sciences of the United States of America 96, 9252-9257.

Robey, R.W., To, K.K.K., Polgar, O., Dohse, M., Fetsch, P., Dean, M. & Bates, S.E. (2009) ABCG2: a perspective. Advanced Drug Delivery Reviews 61, 3-13.

Robinson, G.W., McKnight, R.A., Smith, G.H. & Hennighausen, L. (1995) Mammary epithelial cells undergo secretory differentiation in cycling virgins but require pregnancy for the establishment of terminal differentiation. Development 121, 2079-2090.

Romagnani, P., Annunziato, F., Lazzeri, E., Cosmi, L., Beltrame, C.L., L., Galli, G., Francalanci, M., Manetti, R., Marra, F., Vanini, V., Maggi, E. & Romagnani, S. (2001) Interferon-inducible protein 10, monokine induced by interferon gamma, and interferon-inducible T-cell alpha chemoattractant are produced by thymic epithelial cells and attract T-cell receptor (TCR) alpha beta+ CD8+ single-positive T cells, TCR gamma delta+ T cells, and natural killer-type cells in human thymus. Blood 97, 601-607.

Rudolph, M.C., McManaman, J.L., Hunter, L., Phang, T. & Neville, M.C. (2003) Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. Journal of Mammary Gland Biology and Neoplasia 8(2), 287-307.

Rudolph, M.C., McManaman, J.L., Phang, T.L., Russell, T., Kominsky, D.J., Serkova, N.J., Anderson, S.M. & Neville, M.C. (2007) Metabolic regulation in the lactating mammary gland: a lipid synthesizing machine. Physiological Genomics 28, 323-336.

Rustan, A.C., Nossen, J.O., Christiansen, E.N. & Drevon, C.A. (1988) Eicosapentanoic acid reduces hapatic synthesis and secretion of triacylglycerol by decreasing the activity of

acyl-coenzyme A:1,2-diacylglycerol acyltransferase. Journal of Lipid Research 29, 1417-1426.

S

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. & Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2), 374-378.

Saluz, H.P., Iqbal, J., Limmon, G.V., Ruryk, A. & Wu, Z. (2002) Fundamentals of DNA-chip/array technology for comparative gene-expression analysis. Current Science 83(7), 829-833.

Sauvant, D., Perez, J.M. & Tran, G. (2002) Tables de composition et de valeur nutritive des matières premières destinées aux animaux d'élevage. Editions INRA.

Schennink, A., Heck, J.M.L., Bovenhuis, H., Visker, M.H.P.W., van Valenberg, H.J.F. & Arendonk, J.A.M. (2008) Milk fatty acid unsaturation: genetic parameters and effects of stearoyl-coA desaturase (SCD1) and acyl coA: Diacylglycerol Acyltransferase 1 (DGAT1). Journal of Dairy Science 91, 2135-2143.

Schennink, A., Stoop, W.M., Visker, M.H.P.W., Heck, J.M.L., Bovenhuis, H., Van der Poel, J.J., van Valenberg, H.J.F. & Van Arendonk, J.A.M. (2007) DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Animal Genetics 38, 467-473.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., FGassmann, M., Lightfoot, S., Menzel, W., Granzow, M. & Ragg, T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology 7(3).

Schwertfeger, K.L., McManaman, J.L., Palmer, C.A., Neville, M.C. & Anderson, S.M. (2003) Expression of constitutively activated Akt in the mammary gland leads to excess lipid synthesis during pregnancy and lactation. Journal of Lipid Research 44, 1100-1112.

Scow, R.O., Desnuelle, P. & Verger, R. (1979) Lipolysis and lipid movement in a membrane model. Action of lipoprotein lipase. The Journal of Biological Chemistry 254(14), 6456-6463.

Sheffield, L.G. (1988) Organization and growth of mammary epithelia in the mammary gland fat pad. Journal of Dairy Science 71(10), 2855-2874.

Shuster, R.C., Houdebine, L.M. & Gaye, P. (1976) Studies on the synthesis of casein messenger RNA during pregnancy in the rabbit. European Journal of Biochemistry 71, 193-199.

Sinha, Y.N. & Tucker, H.A. (1969) Mammary development and pituitary prolactin level of heifers from birth through puberty and during the estrous cycle. Journal of Dairy Science 52, 502-512.

Smith, S.J., Cases, S., Jensen, D.R., Chen, H.C., Sande, E., Tow, B., Sanan, D.A., Raber, J., Eckel, R.H. & Farese, R.V. (2000) Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking DGAT. Nature Genetics 25, 87-90.

Sordillo, L.M. & Streicher, K.L. (2002) Mammary gland immunity and mastitis susceptibility. Journal of Mammary Gland Biology and Neoplasia 7, 135-146.

Soukas, A., Cohen, P., Socci, N.D. & Friedman, J.M. (2000) Leptin-specific patterns of gene expression in white adipose tissue. Genes and Development 14, 963-980.

Soyeurt, H., Gillon, A., Vanderick, S., Mayeres, P., Bertozzi, C. & Gengler, N. (2007) Estimation of heritability and genetic correlations for the major fatty acids in bovine milk. Journal of Dairy Science 90, 4435-4442.

Spelman, R.J., Ford, C.A., McElhinney, P., Gregory, G.C. & Snell, R.G. (2002) Characterization of the DGAT1 gene in the New Zealand dairy population. Journal of Dairy Science 83, 3514-3517.

Suchyta, S.P., Sipkovsky, S., Halgren, R.G., Kruska, R., Elftman, M., Weber-Nielsen, M., Vandehaar, M.J., Xiao, L., Tempelman, R.J. & Coussens, P.M. (2003) Bovine mammary gene expression profiling using a cDNA microarray enhanced for mammary-specific transcripts. Physiological Genomics 16, 8-18.

Suzuki, D. (1995) Le gène. In La biologie: Campbell, DeBoeck Université.

Suzuki, N., Chen, N.J., Millar, D.G., Suzuki, S., Horacek, T., Hara, H., Bouchard, D., Nakanishi, K., Penninger, J.M., Ohashi, P.S. & Yeh, W.C. (2003) IL-1 receptor associated kinase 4 is essential for IL-18 mediated NK and Th1 cell responses. The Journal of Immunology 170, 4031-4035.

Swaisgood, H.E. (2003) Chemistry of the caseins. In Advanced Dairy Chemistry - Volume 1 - Proteins - Thrid edition Part A: Fox, P.F. and McSweeney, P.H.L.

Swanson, E.W. & Poffenbarger, J.I. (1979) Mammary gland development of dairy heifers during their first gestation. Journal of Dairy Science 62(702-714).

T

Taniguchi, M., Utsugi, T., Oyama, K., Mannen, H., Kobayashi, M., Tanabe, Y., Ogino, A. & Tsuji, S. (2004) Genotype of stearoyl-CoA desaturase is associated with fatty acid composition in Japanese Black cattle. Mammalian Genome 14, 142-148.

Tremblay, L., Laporte, M.F., Leonil, J., Dupont, D. & Paquin, P. (2003) Quantitation of proteins in milk and milk products. In Advanced dairy chemistry - Volume 1 - Proteins - 3rd edition Part A: Fox, P.F. and McSweeney, P.L.H.

Tucker, H.A. (2000) Hormone, mammary growth, and lactation: a 41-year perspective. Journal of Dairy Science 83, 874-884.

Tusher, V.G., Tibshirani, R. & Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences 98, 4116-4121.

V

van Herwaarden, A.E., Wagenaar, E., Karnekamp, B., Merino, G., Jonker, J.W. & Schinkel, A.H. (2006) Breast cancer resistance protein (Bcrp1/Abcg2) reduces systemic exposure of the dietary carcinogens aflatoxin B1, IQ and Trp-P-1 but also mediates their secretion into breast milk. Carcinogenesis 27(1), 123-130.

Vilotte, J.L., Whitelaw, C.B.A., Ollivier-Bousquet, M. & Shennan, D.B. (2003) Biosynthesis of milk proteins. In Advanced in dairy chemistry - Volume 1 - Proteins - Third edition - Part B, pp. 699-738. Edited by P.F.a.M. Fox, P.L.H.

Vorbach, C., Capecchi, M.R. & Penninger, J.M. (2006) Evolution of the mammary gland from the innate immune system? BioEssays 28, 606-616.

W

Wang, M., Master, S.R. & Chodosh, L.A. (2006) Computational expression deconvolution in a complex mammalian organ. BMC informatics 7, 328.

Ward, A.T., Wittenberg, K.M. & Przybylski, R. (2002) Bovine milk fatty acid profiles produced by feeding diets containing solin, flax and canola. Journal of Dairy Science 85(5), 1191-1196.

Watson, C.J. (2008) Immune cell regulators in mouse mammary development and involution. Journal of Animal Science.

Whittemore, C.T. (1980) Lactation of the dairy cow. Longman Handbooks in Agricultural Series.

Wilgram, G.F. & Kennedy, E.P. (1963) Intracellular Distribution of Some Enzymes Catalyzing Reactions in the Biosynthesis of Complex Lipids. The Journal of Biological Chemistry 238(8), 2615-2619.

Williams, C.M. (2000) Dietary fatty acids and human health. Annales de Zootechnie 49, 165-180.

Winter, A., Alzinger, A. & Fries, R. (2004) Assessement of the gene content of the chromosomal regions flanking bovine DGAT1. Genomics 83, 172-180.

Winter, A., Krämer, W., Werner, F.A.O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J.E., Thaller, G. & Fries, R. (2002) Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at the quantitative trait locus for milk fat content. Proceedings of the National Academy of Sciences of the United States of America 99(14), 9300-9305.

Y

Yamauchi, T., Kamon, J., Waki, H., Terauchi, Y., Kubota, N., Hara, K., Mori, Y., Ide, T., Murakami, K., Tsuboyama-Kasaoka, N., Ezaki, O., Akanuma, Y., Gavrilova, O., Vinson, C., Reitman, M.L., Kagechika, H., Shudo, K., Yoda, M., Nakano, Y., Tobe, K., Nagai, R., Kimura, S., Tomita, M., Froguel, P. & Kadowaki, T. (2001) The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. Nature Medecine 7(8), 941-946.

Yang, L.Y., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. & Speed, T.P. (2002a) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research 30(No. 4 e15), 1-10.

Yang, Y., Dudoit, S., Luu, D., Peng, V., Ngai, J. & Speed, T. (2002b) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research 30(e15).

Annexes

Annexe 1 : Nomenclature des principaux acides gras du lait

Nomenclature	Nom
Acides gras saturés	
C4:0	Acide butyrique
C6:0	Acide caproïque
C8:0	Acide caprylique
C10:0	Acide caprique
C12:0	Acide laurique
C14:0	Acide myristique
C15:0	Acide pentadécanoïque
C16:0	Acide palmitique
C17:0	Acide margarique
C18:0	Acide stéarique
Acides gras monoinsaturés	
C14:1 *cis-9*	Acide myristoléqiue
C16:1 *cis-9*	Acide palmitoléique
C18:1 *cis-9*	Acide oléique
C18:1 *trans-11*	Acide vaccénique
Acides gras polyinsaturés	
C18:2 *cis-9, cis-12*	Acide linoléique
C18:2 *cis-9, trans-11*	Acide ruménique
C18:3 *cis-9, cis-12, cis-15*	Acide linolénique

Annexe 2 : Communications

Polymorphisme au locus DGAT1 et activité transcriptionnelle du tissu mammaire bovin en lactation.*

Faucon, F.[2,3]; K., Duhem[1]; J.C., Mocquot[2]; P., Martin[3]
[1]Centre Interprofessionnel d'Economie Laitière (CNIEL)
[2]Institut de l'Elevage, Paris
[3]Unité Génomique et Physiologie de la Lactation (GPL), Inra, Jouy-en-Josas

Diplomée de l'Ecole Nationale Supérieure Agronomique de Rennes (ENSAR), Félicie FAUCON est spécialisée en Biologie Production et Physico-Chimie du lait. Au cours de son cursus, elle a réalisé deux stages de recherche :
- établissement de corrélation entre analyse sensorielle et analyses physiques comprenant le développement d'un plan expérimental appliqué à l'analyse de la texture des fromages (Aula de Productos Lácteos à Lugo, Espagne)
- étude de l'impact des supplémentations lipidiques de l'alimentation de la vache sur les propriétés physico-chimiques de la matière grasse laitière (MGL) et sur les paramètres zootechniques (Unité Mixte de Recherche sur la Production du Lait, INRA, St-Gilles, en collaboration avec l'UMR Science et Technologie du Lait et de l'Oeuf, INRA, Rennes)

Problématique :

-Observation : baisse de la consommation du lait due à sa teneur élevée en acides gras saturés (AG). Conséquence sur l'économie de toute la filière laitière

-La recherche de QTL (Quantitative Trait Loci) sur les bovins laitiers a permis de découvrir dans la terminaison centromérique du chromosome bovin 14, un gène affectant la quantité de MGL : DGAT1 (Diacylglycerol Acyltransferase 1). La mutation AA en GC entraîne une baisse du taux butyreux (TB)

- Des études sur les chèvres ont mis en évidence un lien étroit entre biosynthèse, transport et sécrétion des micelles de caséines et des globules gras du lait

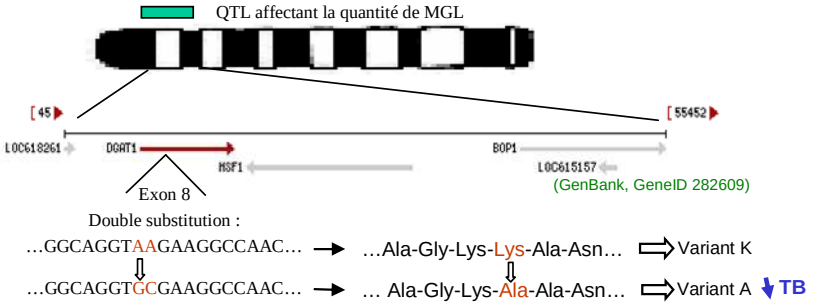


Figure 1 : *QTL, mutation causale et dysfonctionnement d'une protéine, exemple du chromosome bovin 14* (Heyen et al., 1999; Kappes et al., 1999; Grisart et al., 2002; Spelman et al., 2002; Winter et al., 2002 ; Boichard et al., 2003)

Quelles sont les conséquences d'une double substitution au niveau de l'exon 8 du gène DGAT1 sur le fonctionnement de la cellule épithéliale mammaire (CEM) et sur l'organisation de la micelle de caséines et du globules gras?

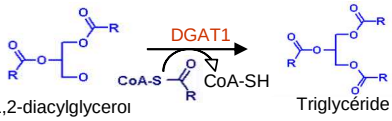


Figure 2 : *l'enzyme DGAT1 catalyse l'étape limitante de la synthèse des triglycérides* (Mayorek et al., 1999 ; Kamoun et al., 2003)

Moyens mis en œuvre :

- **Animaux** : Utilisation du **tissu mammaire** de 6 vaches laitières pleines sœurs en fin de lactation, de génotype contrasté au locus DGAT1. Ces vaches ont été obtenues par transfert embryonnaire à partir des animaux issus du programme de détection des QTL laitiers du Pin au Haras (croisement Holstein*Normande)
- En **transcriptomique** : analyse des différentiels d'expression de la CEM selon le génotype grâce à des **puces à ADN**, puis validation de ces différentiels par **PCRq**
- En **protéomique :** analyse des différentiels d'expression de la CEM selon le génotype par **électrophorèse bidimensionnelle** (E2D) et/ou **chromatographie en phase liquide** (HPLC), puis identification de ces différentiels par **spectrométrie de masse**

* Cette thèse s'insère dans le projet GenoMilkFat (AGENAE), soutenu financièrement par l'ANR et visant à étudier l'impact des facteurs génétiques et alimentaires sur les caractéristiques de la MGL et sur les propriétés techno-fonctionnelles des constituants du lait (micelles de caséines, globules gras, lactosérum). Elle est réalisée dans l'unité GPL de Jouy-en-Josas en convention CIFRE avec l'Institut de l'Elevage.

Institut National de la Recherche Agronomique
30-31 mai 2006
8ème séminaire des thésards du département de Génétique Animale, Limoges, 30-31 mai 2006

Gènes impliqués dans le développement et la différenciation fonctionnelle de l'épithélium sécrétoire mammaire caprin

[1, 2]Faucon, F., [2]Zalachas, E., [3]Robin, S. et [2]Martin, P.

[1] Institut de l'Elevage, Paris

[2] Unité Génomique et Physiologie de la Lactation (GPL), Inra, Jouy-en-Josas

[3] Unité Mixte de Recherche 518 du département Mathématiques et Informatique Appliquées (MIA), AgroParisTech, Paris

Le développement de l'épithélium sécrétoire mammaire et sa différenciation fonctionnelle s'opèrent au cours de la gestation sous l'effet combiné de stéroïdes ovariens, d'hormones peptidiques hypophysaires et de facteurs de croissance. Si l'effet de ces molécules est (relativement) bien connu, l'effet de facteurs de différenciation exprimés localement reste, en revanche, assez peu documenté. Des résultats antérieurs obtenus à partir d'une comparaison de profils transcriptionnels mammaires de chèvres, à 70 et 110 jours de gestation (banques SAGE, microarrays, RT-PCR quantitative en temps réel), ont révélé un fort différentiel d'expression du gène spécifiant la caséine κ. L'expression des caséines ayant été rapportée chez plusieurs espèces (rongeurs notamment) au cours de la gestation, ce résultat n'est pas surprenant en soi. C'est d'avantage par le caractère spécifique (caséine κ) et l'ampleur du phénomène qu'il est remarquable.

Pour apporter des éléments de réponse à cette observation une analyse transcriptomique étendue à 5 stades physiologiques (4 stades de gestation, G45, G70, G90, G110 et lactation) a été entreprise. Un plan d'expérience approprié a été mis en place pour réaliser un suivi longitudinal des profils d'expression génique au cours de la différenciation du tissu mammaire. Auparavant, la faisabilité de la démarche a été évaluée en comparant le stade le plus précoce de la gestation (G45) à la lactation. La comparaison de ces deux stades de développement s'est faite sur trois "dye-swap" en confondant l'effet « individu » avec la réplication technique, et en utilisant des lames sur lesquelles a été spotté le set d'oligonucléotides bovin Operon. L'analyse statistique a fait apparaître que sur le référentiel de 8 379 gènes spottés, 6 923 peuvent être analysés sans ambiguïté (83%). Parmi les gènes exprimés, 4% le sont de façon équivalente dans les deux tissus. Après correction statistique de Benjamini-Hochberg, 299 gènes ont été déclarés différemment exprimés entre le début de la gestation (45 jours) et la lactation. L'interrogation des bases de données montre que le tissu en gestation surexprime des gènes correspondant à l'activité des cellules du tissu conjonctif. En revanche, le tissu mammaire lactant surexprime des gènes connus pour être impliqués dans les voies de biosynthèse des protéines et des acides gras du lait.

Journées ABIES – 21 ,22 mars 2007

Effets du polymorphisme au locus DGAT1 sur l'activité transcriptionnelle du tissu mammaire bovin au cours de la lactation - Impact sur la composition et sur la qualité nutritionnelle des fractions protéiques et lipidiques du lait.

Résumé :
La filière laitière, secteur clé de l'industrie agro-alimentaire, est aujourd'hui confrontée à des remises en questions majeures. Le lait et particulièrement sa fraction lipidique, du fait de sa richesse en acides gras (AG) saturés (62 % du poids total des acides gras du lait d'après Precht et al., 1999) supposés athérogènes lorsqu'ils sont surconsommés (Legrand et al., 2001; Bingham, 2002), est la cible d'attaques répétées et connaît une désaffection des consommateurs dont les répercussions économiques sont considérables. Pour remédier à cette situation, la stratégie envisagée par les acteurs de la filière est de développer une politique d'acquisition systémique de connaissances sur le lait et sa matière grasse de façon à proposer un ensemble cohérent d'outils qui permettront d'adapter le profil en AG aux recommandations nutritionnelles, en exploitant une plasticité de composition obtenue sous l'effet de facteurs génétiques et de certaines pratiques d'élevage (alimentation, notamment), qui peuvent éventuellement être combinés.
L'étude approfondie de la synthèse du lait décrit des interactions mécanistiques fortes entre voies de biosynthèse, d'assemblage et de sécrétion des constituants majeurs du lait (globules gras et micelles de caséines, en particulier) (Mather et Keenan, 1998). Afin de mieux comprendre ces interactions, le projet de thèse vise à préciser l'impact de la mutation K232A au locus DGAT1 affectant l'activité de l'enzyme spécifiée par ce gène sur la composition du lait, au travers d'une analyse fonctionnelle du tissu mammaire.
Pour ce faire, le programme consistera d'abord à générer un corpus de données informatives en procédant à une analyse expressionnelle (transcriptome et protéome à partir des techniques de puces à ADN, de PCR quantitative en temps réel et d'électrophorèse) du tissu mammaire.
Cette étude s'inscrit dans le cadre d'un programme plus large (Geno Milk Fat) soutenu par l'ANR.

1er colloque Doc'J, Jouy-en-Josas, 22-23 mai 2006

58th Annual Meeting of the European Association for Animal Production, Dublin, Ireland, 26-29 August 2007

Session 34, theatre 8, p.347 of the book 13.

Genes involved in development and functional differentiation of caprine mammary secretory epithelium

F. Faucon[1,2] , E. Zalachas[2] , S. Robin[3] and P. Martin[2] , [1]Institut Elevage, Paris, 75012, France , [2]Unité Génomique et Physiologie de la Lactation, INRA, Jouy en Josas, 78352, France , [3]Unité Mixte de Recherche 518, AgroParisTech, Paris, 75005, France

Development of mammary secretory epithelium and its functional differentiation occur during pregnancy under combined actions of ovarian steroids, pituitary hormones and growth factors. If the effect of these molecules is relatively well known, effect of differentiation factors expressed locally is not enough characterized. To understand local regulation of mammary tissue development and differentiation we realized transcriptional analysis on 5 physiological stages (4 during pregnancy and 1 during lactation). An appropriate experimental design was drawn to follow gene expression profiles during differentiation of mammary tissue. We first validated experimental feasibility by comparing expression pattern of mammary tissue from the first pregnancy (45 days) and from lactation stages. Using 3 goats per stage, this comparison was done on 3 dye-swaps, utilizing oligoarrays. Statistical analysis revealed that among the 8 379 genes spotted, 299 genes were differentially expressed between pregnancy and lactation. Data bases examination confirmed that genes controlling cellular activity of conjunctive tissue were over-expressed during pregnancy while genes implicated in milk protein and fatty acid biosynthesis were over-expressed during lactation.

Différenciation fonctionnelle de l'épithélium sécrétoire mammaire caprin
Functional differentiation of caprine mammary secretory epithelium

FAUCON F. (1, 2), ZALACHAS E. (2, 4), ROBIN S. (3), MARTIN-MAGNIETTE M.L. (3) et MARTIN P. (2, 4)
(1) Institut de l'Elevage, Paris
(2) Unité Génomique et Physiologie de la Lactation (GPL), Inra, Jouy-en-Josas
(3) Unité Mixte de Recherche 518 du département Mathématiques et Informatique Appliquées (MIA), AgroParisTech, Paris
(4) Plateforme d'Instrumentation et de Compétence en Trancriptomique (PICT), Inra, Jouy-en-Josas

INTRODUCTION

Le développement de l'épithélium sécrétoire mammaire et sa différenciation fonctionnelle s'opèrent au cours de la gestation sous l'effet combiné de stéroïdes ovariens, d'hormones hypophysaires et de facteurs de croissance (Gilbert, 2006). Si l'effet de ces molécules est relativement bien connu, l'effet de facteurs de différenciation exprimés localement reste, en revanche, assez peu documenté (Henninghausen et Robinson, 1998). Pour progresser dans la connaissance de ces mécanismes, une analyse du trancriptome étendue à 5 stades physiologiques a été entreprise.

1. MATERIEL ET METHODES

1.1. ANIMAUX

Un échantillon de tissu mammaire de douze chèvres gestantes (Alpine et Saanen) réparties sur quatre stades de gestation (45 jours, G45 ; 70 jours, G70 ; 90 jours, G90 et 110 jours, G110) et de trois chèvres primipares à 40 jours de lactation (L40) a été prélevé après abattage des animaux. Les tissus ont été immédiatement lavés dans une solution PBS avant d'être congelés dans l'azote liquide et conservés à -80°C.

1.2. TECHNIQUE DES PUCES A ADN

Les ARN totaux ont été extraits à partir des échantillons de tissu mammaire en utilisant le TRIzol Reagent (Invitrogen Life Technologies) puis traités à la DNase sur colonnes Cleanup (Qiagen) afin d'éliminer l'ADN génomique. La quantité et la pureté des ARN extraits ont été évaluées à partir des mesures d'absorbance sur NanoDrop ND-100 (Nixor Biotech) à 260, 280 et 230 nm. Leur intégrité a été vérifiée sur Bioanalyseur 2100 en utilisant le kit RNA 6000 nano LabChip (Agilent Technologies). Seuls les ARN dont le *RNA Integrity Number* (RIN) était supérieur à 7 ont été retenus. Les ARN totaux de chaque échantillon de tissu mammaire (10µg) ont été rétro-transcrits et marqués avec des fluorochromes (cy3 et cy5, Amersham Biosciences Europe) en utilisant le kit SuperSript Indirect cDNA Labeling System (Invitrogen Life Technologies) selon les instructions du fabricant. La quantité des cibles marquées et purifiées et l'efficacité des marquages ont été déterminées après mesure des absorbances à 260, 550 et 650 nm. La qualité des marquages a été évaluée sur gel d'agarose à 1%. 380ng de chaque cible marquée (correspondant à 40 pmol de cyanine) ont été hybridés sur 18 lames de verre selon un plan expérimental en boucle (Kerr et Churchill, 2001). Les sets d'oligonucléotides bovin Illumina et Operon on été déposés sur les microarrays utilisés (CRB, INRA, France).

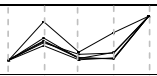
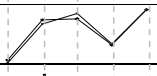
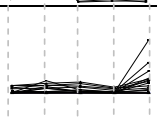
1.3. EXTRACTION ET ANALYSE DES DONNEES

Les lames ont été scannées à l'aide d'un scanner Agilent. Les images ont été quantifiées avec le logiciel Feature Extraction (Agilent Technologies). A partir des données brutes extraites de l'analyse des huit premières lames, les profils d'expression des gènes des protéines du lait et de la lipogenèse ont été tracés.

2. RESULTATS

Quatre profils ont été obtenus (cf. tableau 1). Le profil 1 contient des gènes dont le signal est plus intense à G70 et L40. Ce profil caractérise l'essentiel des protéines du lait à l'exception de la caséine beta (profil 2) et de la beta-lactoglobuline absente du répertoire. Le profil 3 caractérisé par une forte expression à G70 est observé pour une quinzaine de protéines dont la lactoferrine et l'ostéopontine. Le profil 4 correspond à un ensemble de gènes dont l'expression s'accroît considérablement pendant la lactation. Parmi ceux-ci on peut noter un nombre significatif d'enzymes de la lipogenèse.
Ces profils ont été validés par PCR quantitative en temps réel pour les protéines indiquées dans le tableau 1.

Tableau 1 Profil d'évolution de la quantité des transcrits au cours du développement de la glande mammaire

N°	Graphe	Gènes associés
	G46 G70 G90 G110 L40	
1		Caséine alpha S1, caséine alpha S2, alpha-lactalbumine, caséine kappa
2		Caséine beta
3		Lactoferrine, ostéopontine
4		Acétyl-CoA carboxylase (ACC), Adipophiline (ADRP), butyrophiline, DGAT1, FABP, Stearoyl Coenzyme A desaturase (SCD), lipoprotein lipase (LPL), Fatty acid synthase (FAS)

3. DISCUSSION

La présence de transcrits de gènes codant pour des protéines du lait à 70 jours de gestation chez la chèvre suggère que la cellule épithéliale mammaire a atteint un stade de différenciation suffisant pour synthétiser les constituants du lait. Cette observation a déjà été faite chez la souris en gestation.
Par ailleurs, il a été observé dans le tissu mammaire caprin à 70 jours de gestation la présence de protéines du lait (données non publiées). La différenciation s'étendrait donc jusqu'au processus de traduction.

CONCLUSION

D'autres expériences en morphologie et en protéomique sont actuellement en cours. Leurs résultats seront utilisés une fois l'ensemble des 18 lames analysées afin de dégager une vision globale et intégrée de la différenciation du tissu mammaire chez la chèvre gestante.

Gilbert, S.F., 2006. Development Biology, chap. 17
Hennighausen, L., Robinson, G.W., 1998. Genes Dev., 12, 449-445
Kerr, M.K., Churchill, G.A., 2001. Biostatistics, 2, 183-201

EXPRESSION DE GENES DE L'IMMUNITE DANS LE TISSU MAMMAIRE CAPRIN GESTANT

FAUCON F.[1,2], REBOURS E.[1], HELBLING J.-C.[1], BEVILACQUA C.[1], AUBERT J.[3], MARTIN-MAGNIETTE M.L.[3], ROBIN S.[3] et MARTIN P.[1]

[1]*Unité Génomique et Physiologie de la Lactation (GPL), équipe GEL-PICT, Inra, Jouy-en-Josas*
[2]*Institut de l'Elevage, Paris*
[3]*Unité Mixte de Recherche 518, AgroParisTech, Paris*
felicie.faucon@jouy.inra.fr

INTRODUCTION

Le développement du tissu mammaire et sa différenciation fonctionnelle s'opèrent au cours de la gestation sous l'effet combiné de stéroïdes ovariens, d'hormones hypophysaires et de facteurs de croissance (Gilbert, 2006). Si l'effet de ces molécules est relativement bien connu, l'effet de facteurs de différenciation exprimés localement reste, en revanche, assez peu documenté (Henninghausen et Robinson, 1998). Pour progresser dans la connaissance de ces mécanismes, nous avons entrepris d'analyser le transcriptome du tissu mammaire, à différents stades de gestation et en lactation.

MATERIEL ET METHODES

Animaux

Des prélèvements de tissu mammaire ont été effectués, sur douze chèvres gestantes à 4 stades de gestation (46 jours, G46 ; 70 jours, G70 ; 90 jours, G90 et 110 jours, G110) et sur trois chèvres primipares, à 40 jours de lactation (L40). Les échantillons de tissus prélevés ont été conservés à -80°C.

Analyse des profils d'expression

Les ARN totaux, extraits (TRIzol Reagent, Invitrogen Life Technologies) ont été traités à la DNase sur colonnes Cleanup (Qiagen). 10µg d'ARNt ont été retro-transcrits et marqués (Cy3 et Cy5, Amersham Biosciences Europe) selon le kit Superscript™ Indirected cDNA labelling System (Invitrogen Life Technologies). La quantité et la qualité des ARNt et des cibles marquées ont été contrôlées (NanoDrop ND-100, Nixor Biotech et Bioanalyzer 2100, Agilent Technologies). Les cibles marquées (380ng) ont été hybridées par paires sur lames (18) selon un plan expérimental en boucle (Kerr & Churchill, 2001). Ces lames fournies par le Centre de Ressources Biologiques GADIE (INRA, Jouy-en-Josas), contenaient 22 000 sondes géniques bovines. Les données extraites (scanner, logiciel FeatureExtraction, Agilent Technologies) ont été normalisées (Anapuce) puis analysées par régression linéaire afin d'établir le profil d'expression de chaque gène. Les gènes de profil semblable ont été regroupés (fonction KMC, logiciel TMeV) et les fonctions biologiques des groupes identifiées à l'aide du logiciel Ingenuity Pathway Analysis (IPA) (Ingenuity Systems)

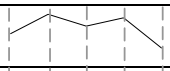
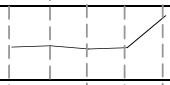
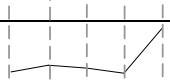
PCR quantitative en temps réel

Les systèmes spécifiques des gènes analysés ont été conçus avec le logiciel PrimerExpress (Applied Biosystems) et les analyses (SYBR Green PCR MasterMix) effectuées sur un appareil SDS 7900 Applied Biosystems.

RESULTATS

Trois profils ont été plus particulièrement analysés (Tableau 1). Le premier correspond à des gènes présentant un signal plus intense à G70 et G110 qui interviennent dans la réponse immunitaire. Le profil 2, représente des gènes plus exprimés en lactation et participant au métabolisme lipidique. Le profil 3 correspond à des gènes dont l'expression s'accroît sensiblement à G70 et L40. Ces gènes codent des protéines impliquées dans le métabolisme des carbohydrates (dont le lactose).

Tableau 1 Evolution de la quantité des transcrits au cours du développement et de la différenciation de la glande mammaire

N°	Profils	Fonctions biologiques IPA
	G46 G70 G90 G110 L40	
1		Réponse immunitaire et immunité
2		Métabolisme lipidique, transport moléculaire
3		Métabolisme des carbohydrates et des acides aminés

DISCUSSION/CONCLUSION

L'expression des gènes du métabolisme lipidique et de la synthèse du lactose en lactation confirme que la différenciation totale du tissu mammaire n'intervient qu'après la parturition, les gènes de la lipogenèse et de la synthèse du lactose ne s'exprimant que dans le tissu mammaire en lactation comme cela avait été décrits chez la souris ou le lapin (Anderson et al., 2007). Cette étude met en évidence la présence de transcrits spécifiant des molécules impliqués dans la réponse immunitaire en milieu de gestation. Cette observation suggère que le tissu mammaire aurait déjà acquis à cette étape les propriétés d'un épithélium avec un système immunitaire fonctionnel, tel que décrit par Sordillo et Streicher en 2002. Ceci va dans le même sens que la constatation de transcrits des lactoprotéines dans la glande mammaire de souris gestantes (Robinson et al., 1995).

Ainsi, le tissu mammaire aurait atteint un certain degré de différenciation en milieu de gestation.

Anderson, S.M., et al., 2007. Breast Cancer Res, 9, 204-218
Gilbert, S.F., 2006. Development Biology, chap. 17
Hennighausen, L., Robinson, G., 1998. Genes Dev., 12, 449-55
Kerr, M.K., Churchill, G.A., 2001. Biostatistics, 2, 183-201
Robinson, G.W. et al., 1995. Development, 121, 2079-2090
Sordillo, L.M., Streicher, K.L., 2002. J Mammary Gland Biol Neoplasia, 7, 135-146

Journées Scientifiques du département de Génétique Animale INRA, Lacanau, 21-24 avril 2008, p.31

Title:…Goat mammary gland expressed genes representative of immunity function during pregnancy before the acquisition of its secretory phenotype

Authors;:
Presenting Author (name & address): FAUCON Félicie Unité GPL Bâtiment 221 INRA Jouy en Josas Domaine de Vilvert 78352 Jouy en Josas
.
Conference registration number of Presenting Author: 1
Other authors (name only):
1. REBOURS E.
2. HELBLING J.C.
3. BEVILACQUA C.
4. AUBERT J.
5. MARTIN-MAGNIETTE M.L.
6. ROBIN S.
7. MARTIN P.

Abstract (maximum 200 words):
Mammary tissue differentiation into a secretory epithelium that synthesizes milk occurs during pregnancy. Impaired differentiation during pregnancy may provoke lactating default. Molecular mechanisms associated to mammary tissue terminal differentiation remained incompletely understood especially in ruminants. Morphological differences between mice and ruminant mammary tissues before pregnancy let's think that there are differences between these two mammals in the differentiation mechanisms. A transcriptional analysis on 5 physiological stages (4 during pregnancy and 1 during lactation) appeared as the best approach to get overview of the differentiation process. An appropriate experimental design was drawn to follow gene expression profiles during differentiation of mammary tissue. Using 3 nulliparous or primiparous goats per stage (a small ruminant model economically acceptable), the comparison was done on 9 dye-swaps, using 22k bovine oligoarrays. Statistical analysis revealed that 2 330 genes varied significantly at least once. These genes were divided into 19 clusters. Identification of biological functions revealed that mechanisms described for mice can be applied to goats. As for example the fall down in lactation of genes implicated in cell growth and proliferation and the increase in lactation of expression of genes implicated in lipid metabolism. A special observation, not described for mice, was the expression of immunity genes at mid-pregnancy.

XXXI Conference of the International Society for Animal Genetics, 20-24 juillet 2008, Amsterdam, Pays-Bas

Bourse de voyage pour les étudiants
Poster primé

Author(s): FAUCON F.[1,2], REBOURS E.[1], BEVILACQUA C.[1], HELBLING J.-C.[1], AUBERT J.[3], MAKHZAMI S.[1], MARTIN-MAGNIETTE M.L.[3], ROBIN S. [3] and MARTIN P.[1]

[1] INRA, UR1196 Génomique et Physiologie de la Lactation, équipe GEL-PICT, F-78352 Jouy-en-Josas
[2] Institut de l'Elevage, F-75012 Paris
[3] UMR INAPG/ENGREF/INRA 518, Mathématiques et Informatique Appliquées, F-75005 Paris

Presentation Title:

During pregnancy, the goat mammary tissue first express genes representative of the immune function before the acquisition of its complete secretory phenotype

Abstract:

Mammary tissue terminal differentiation into a functional epithelium that synthesizes and secretes milk occurs during pregnancy under the combined effects of ovarian steroids, pituitary peptide hormones and growth factors. Molecular mechanisms underlying this complex process remain incompletely understood, especially for ruminants. Although several mammary genes have been identified so far, it is likely that locally expressed factors still have to be found. A time-course transcriptional analysis, performed with microarrays, on 5 physiological stages (4 during pregnancy and 1 during lactation) appeared as a valuable approach to get an overview of the ruminant mammary gland differentiation process. Goats were chosen for this analysis as they are small ruminant model economically acceptable. An appropriate loop experimental design was drawn to follow gene expression profiles during terminal differentiation of the mammary tissue. Using 3 nulliparous (pregnancy) or primiparous (lactation) goats per stage, a comparison was done on 9 dye-swaps, using 22K bovine oligoarrays (CRB-GADIE, INRA, France). Statistical analysis revealed that 2 330 genes varied significantly at least once in the study. These genes were divided into 19 clusters according to their expression profile. Identification of biological functions revealed that general mechanisms well described for mice can be applied to goats. As for example, the expression at mid-pregnancy of genes encoding some milk proteins, the downfall after parturition of genes involved in cell growth and proliferation and the increase in expression, with lactation, of genes implicated in lipid metabolism. A remarkable observation, not underlined so far in pregnant mice mammary tissue, was the identification of immune system-related genes at mid-pregnancy, in the goat mammary tissue. This observation suggests that the first half of pregnancy is crucial to obtain a healthy and productive mammary gland and that a special breeding attention has to be paid during this period. Indeed, mid-pregnant goat mammary tissue already express genes encoding milk proteins as well as a large repertoire of immune-related genes to ensure later on an effective mucosal protection.

5ème Symposium IMGC – Sydney, 14-16 octobre 2008

Bourse de voyage pour les étudiants

Effet du polymorphisme (AA/GC) au locus *DGAT1* sur l'activité transcriptionnelle du tissu mammaire bovin, sur la composition du lait et sur les caractéristiques des globules gras et des micelles de caséines

FAUCON F. (1,2), REBOURS E. (1), BERNARD L. (3), HURTAUD C. (4), MIRANDA G. (1), MENARD O. (5), DHORNE-POLLET S. (1), BEVILACQUA C. (1), LARROQUE H. (6), GALLARD Y. (7), LEROUX C. (3), MARTIN P.(1)
(1) INRA, UR1196 Génomique et Physiologie de la Lactation, équipe GEL-PICT, F-78352 Jouy-en-Josas
(2) Institut de l'Elevage, F-75012 Paris
(3) INRA, UR1213 Herbivores, équipe TALL, F-63122 Saint-Genès-Champanelle
(4) INRA, AgroCampus Ouest, UMR1080 Production du lait, F-35590 Saint-Gilles
(5) INRA, AgroCampus Ouest, UMR1253 Science et technologie du lait et de l'œuf, F-35042 Rennes
(6) INRA, UR337 Station de génétique quantitative et appliquée, F-78352 Jouy-en-Josas
(7) INRA, UE326 Unité expérimentale du Pin-au-Haras, F-61310 Exmes

RESUME
La recherche de QTL laitiers bovins affectant la matière grasse laitière (MGL) a conduit à l'identification d'un polymorphisme (AA vs GC) au niveau du gène spécifiant l'enzyme Diacylglycérol Acyltransferase 1 (DGAT1) impliquée dans l'étape ultime et limitante de la synthèse des triglycérides, qui composent majoritairement (97%) la MGL. Ce polymorphisme modifie la séquence en acides aminés de la protéine (K232A), ce qui entraîne une réduction de la vitesse maximale de la réaction catalysée par DGAT1 et donc de son efficacité. Notre objectif était d'étudier l'effet de ce polymorphisme : d'une part sur la composition et les caractéristiques des structures supramoléculaires du lait (globules gras et micelles de caséines), et d'autre part sur le transcriptome de la région sécrétrice du tissu mammaire. Pour cela, 7 couples de vaches laitières pleine-sœurs ou demi-sœurs (même père) et de génotypes opposés (homozygotes AA/AA et GC/GC) au locus *DGAT1* ont été choisis au sein du dispositif de détection de QTL laitiers (croisement Holstein x Normande) mis en place sur l'installation expérimentale du Pin-au-Haras. La taille et la composition des globules gras et des micelles de caséines des laits ont été analysées. Une étude transcriptomique a été réalisée à partir des ARN totaux extraits de biopsies de tissu mammaire et hybridés sur des lames bovines 22K produites par le CRB de Jouy-en-Josas et des lames bovines 4x44K d'Agilent Technologies. L'analyse des laits confirme une diminution du taux butyreux (TB) et montre une réduction de la taille des globules gras pour les individus de génotype GC/GC (A232), associé à un TB faible. La composition en acides gras des laits varie dans le sens d'une augmentation de l'acide myristique (C14 :0) et des acides gras longs et insaturés (C18 :1 et C20 :1) chez des individus homozygotes pour l'allèle GC (A232). L'analyse transcriptomique a mis en évidence 1 496 gènes différemment exprimés entre les deux génotypes au locus *DGAT1*. Ces gènes sont principalement impliqués dans les voies du métabolisme glucidique et lipidique.

Effect of the polymorphism (AA/GC) occurring at the *DGAT1* locus on the transcriptional activity of the bovine mammary tissue, on the milk content and on the milk fat globules and casein micelles characteristics

FAUCON F. (1,2), REBOURS E. (1), BERNARD L. (3), HURTAUD C. (4), MIRANDA G. (1), MENARD O. (5), DHORNE-POLLET S. (1), BEVILACQUA C. (1), LARROQUE H. (6), GALLARD Y. (7), LEROUX C. (3), MARTIN P.(1)
(1) INRA, UR1196 Génomique et Physiologie de la Lactation, équipe GEL-PICT, F-78352 Jouy-en-Josas
(2) Institut de l'Elevage, F-75012 Paris

SUMMARY
Bovine milk QTL research revealed a polymorphism (AA vs GC) in the gene encoding the Diacylglycerol Acyltransferase 1 (DGAT1) enzyme. This enzyme catalyzes the ultimate and limiting step of the triacylglyceride (TAG) synthesis which represent the major milk fat fraction (97%). This polymorphism modifies the amino acid sequence of the protein (K232A) which affects the maximum velocity (Vmax) and the efficiency of the enzyme. The aim of this work was to study the effect of the K232A polymorphism on (i) the milk composition and milk fat globule and casein micelle characteristics, (ii) the transcriptome of the mammary tissue. Seven paired cows full-sib or half sib (same father) with opposite and homozygous genotype at the locus *DGAT1* were chosen from the milk QTL herd reared in an experimental farm at Le Pin-au-Haras (INRA, France). Milk samples were analysed for fat globule and casein micelle size and composition. Transcriptomic study was performed on total RNA extracted from mammary biopsies and hybridized on 22K bovine slides produced at CRB in Jouy-en-Josas (INRA, France) and 4x44K bovine slides from Agilent Technologies. Milk fat content and the fat globule size decreased whereas myristic (C14:0) as well as long and unsaturated fatty acid (C18:1 and C20:1) content increased for individuals of genotype GC/GC A232 at the *DGAT1* locus. From the transciptomic analysis, 1 496 genes were shown to be differentially expressed between the two DGAT1 genotypes. These genes are mainly involved in glucose and lipid metabolism pathways.

INTRODUCTION

La filière laitière connaît actuellement une phase critique en partie due à la baisse de consommation du lait liée à la teneur élevée en acides gras saturés de sa matière grasse (Matière Grasse Laitière, MGL), essentiellement constituée de triglycérides (97%). Les acides gras saturés étant reconnus athérogènes lorsqu'ils sont consommés de façon excessive (Bingham, 2002). L'enzyme qui catalyse l'étape ultime et limitante de la synthèse des triglycérides, est la Diacylglycérol Acyltransferase 1. Il a été montré chez la souris que l'invalidation du gène spécifiant cette enzyme conduit à une absence totale de production laitière (Smith et al., 2000). Chez les bovins laitiers, le locus *DGAT1* est localisé dans la terminaison centromérique du chromosome

14 (BTA14) où un QTL affectant la quantité de lait, de MGL produite et la concentration de MGL a été détecté (Coppieters et al., 1998 ; Boichard et al., 2003). La différence de quantité de MGL produite, mais également de taux butyreux (TB), entre individus est pour partie expliquée par la substitution de deux nucléotides successifs (AA → GC) localisée dans le huitième exon du gène DGAT1 et conduisant aux deux formes allèliques du gène les plus répandues (Grisart et al., 2002). Cette double mutation induit une substitution Lysine/Alanine en position 232 de la chaîne peptidique (K232A) de la protéine qui a pour effet de modifier la vitesse maximale de l'enzyme (Grisart et al., 2004) et l'affinité pour son substrat, expliquant pour partie le changement de composition en acides gras (AG) du lait observé par Schennink et al. (2007).

Notre objectif était d'une part de valider l'effet de cette double mutation au locus *DGAT1* sur les performances laitières et sur la composition en acides gras du lait des animaux choisis à partir du dispositif mis en place au Pin-au-Haras (croisement Holstein x Normande, Larroque et al., 2003) et d'autre part d'en étudier l'effet sur la taille des globules gras, des micelles de caséines et sur l'activité transcriptionnelle du tissu mammaire, afin de mieux cerner les voies métaboliques affectées par ce polymorphisme.

Pour plus de clarté, le génotype AA/AA (K232) au locus *DGAT1* sera appelé génotype « fort » (pour le critère TB) et le génotype GC/GC (A232), génotype « faible ».

1. MATERIELS ET METHODES

1.1. ANIMAUX

Quatorze vaches laitières homozygotes au locus *DGAT1* issues de trois pères différents ont été sélectionnées au sein du dispositif de détection de QTL laitiers du Pin-au-Haras. A chaque vache de génotype homozygote fort (AA/AA) au locus *DGAT1* était associé, soit sa pleine sœur (n=4), soit sa demi-sœur (même père, n=3) de génotype homozygote faible (GC/GC) , pour constituer *in fine* 7 couples de vaches de génotypes opposés. En 2007, année au cours de laquelle fut mis en place le protocole à l'INRA de Theix (Installations Expérimentales de l'URH), deux couples de vaches étaient en 3ème lactation (L3), deux couples en 2ème lactation (L2) et trois couples en 1ère lactation (L1). L'originalité de ce dispositif par rapport aux études précédemment réalisées sur la mutation K232A au locus *DGAT1* réside dans le fait que les vaches d'un même couple étaient de même stade de lactation, étaient homozygotes et de génotype opposé au locus *DGAT1*, se trouvaient dans le même élevage et recevaient le même régime alimentaire au moment des prélèvements.

1.2. MESURES PHENOTYPIQUES

A 230 jours (± 20 jours), les performances laitières ont été mesurées sur une semaine, alors que les animaux recevaient *ad libitum* un régime de type hivernal à base d'ensilage de maïs. Les variables mesurées étaient : la production laitière journalière (lait en kg/j), les taux butyreux (TB) et protéique (TP) et la teneur en lactose (TL en g/kg), la quantité de cellules dans le lait (en millier/ml), le poids de l'animal (en kg). A partir de ces valeurs mesurées, étaient calculées les quantités de matière grasse (MG = TB x lait), protéique (MP = TP x lait) et de lactose (ML = TL x lait en g/j) secrétées. La composition du lait en acides gras et la taille des globules gras ont été mesurées sur un échantillon de lait de la traite du matin. La composition en protéines majeures et la taille des micelles de caséine ont été évaluées sur un échantillon de lait de la traite du soir. Les performances zootechniques et les mesures sur les laits, ont été analysées statistiquement (analyse de variance) en utilisant la fonction *lme* (modèle linéaire mixte) du package « nlme » du logiciel R. Le modèle d'analyse de variance était :

y = gen + lact + gen x lact + *Père* + e,

Avec

y : variable quantitative étudiée ;

« gen » : effet fixe du polymorphisme au locus *DGAT1* ;

« lact » : effet fixe du stade de lactation ;

Père : effet aléatoire du père ;

« gen x lact » : interaction entre l'effet « gen » et l'effet « lact » ;

e : résiduelle du modèle.

1.3. ANALYSE TRANSCRIPTOMIQUE

Une biopsie mammaire a été réalisée après anesthésie locale de la mamelle, suivant la méthode de Farr et al. (1996). Les ARN totaux (ARNt), extraits des biopsies (TRIzol Reagent, Invitrogen, Life Technologies) ont été traités à la DNase sur colonnes Cleanup (Qiagen). Les ARNt (1 µg) ont été retro-transcrits, puis marqués (Cy3 et Cy5) au moyen du kit Low RNA Linear Amplification Kit PLUS, Two-Color (Agilent Technologies). La quantité et la qualité des ARNt et des cibles marquées ont été contrôlées (NanoDrop ND-100, Nixor Biotech et Bioanalyzer 2100, Agilent Technologies). Les cibles marquées ont ensuite été hybridées, par couple de vaches (cf. §1.1.), en dye-swap sur lames commerciales Agilent bovines 4 x 44K (825 ng déposés) et sur lames bovines 22K produites par le Centre de Ressources Biologiques (CRB)-GADIE (INRA, Jouy-en-Josas) à raison de 2,7 à 6 µg déposés par cible. Les données extraites (scanner, logiciel Feature Extraction, Agilent Technologies) ont été normalisées (fonction *normalisation*, package « anapuce », logiciel R). Après vérification de leur cohérence, les deux jeux de données issus des deux types de lames ont été regroupés afin d'obtenir un nouveau jeux de données comparable à une « super lame » contenant 43 061 cibles (soit 16 091 gènes différents possédant un identifiant Ingenuity). Ce nouveau jeux a été vérifié puis analysé par analyse différentielle (*DiffAnalysis*, « anapuce », R) en regroupant les gènes de même variance (Delmar et al., 2005) et en corrigeant les faux positifs par la méthode de Benjamini et Hochberg (1995). Les fonctions métaboliques associées à la liste de gènes différemment exprimés entre les deux génotypes au locus *DGAT1* ont été identifiées à l'aide du logiciel Ingenuity Pathway Analysis (IPA).

2. RESULTATS

2.1. EFFET DU POLYMORPHISME AU LOCUS *DGAT1* SUR LES PERFORMANCES LAITIERES ET LES CARACTERISTIQUES PHYSICO-CHIMIQUES DU LAIT

2.1.1. Réduction du TB et de la taille des globules gras

Le tableau 1 présente l'effet du polymorphisme au locus *DGAT1* sur les performances laitières et sur les caractéristiques des laits. Les animaux de génotype « faible » présentent un TB et une taille des globules gras significativement plus faible que les animaux de génotype « fort » (-8,6 g/kg et -0,7 µm, respectivement). Le génotype au locus *DGAT1* n'a en revanche pas d'effet significatif sur la production de lait, le TP, les matières sécrétées (MG et

MP), le taux de cellules et sur la taille des micelles de caséines. De même, ce polymorphisme n'a pas d'effet significatif sur la teneur en lactose et la quantité de lactose sécrétée (ML) malgré une tendance à une légère élévation avec le génotype « faible ».

Tableau 1 : Effets du génotype au locus *DGAT1* sur les performances zootechniques, la taille des globules gras et des micelles de caséines (ML : quantité de lactose secrétée, MG : matière grasse, MP : matière protéique, NS : effet non significatif à 5%, TB : taux butyreux, TGG : taille globules gras, TL : teneur en lactose, TMC : taille micelles de caséines, TP : taux protéique, * effet de l'interaction génotype x lactation)

	AA/AA (K232)	GC/GC (A232)	P-value
lait (kg/j)	16,3	18,9	NS
TB (g/kg)	**50,9**	**42,3**	**P<0,01**
TP (g/kg)	34,4	32,8	NS*
TL (g/kg)	47,3	49,8	NS
Cellules (millier/ml)	266,1	291,6	NS*
MG (g/j)	828,6	789,5	NS*
MP (g/j)	563,2	615,9	NS*
ML (g/j)	773,8	932,9	NS
TGG (µm)	**5,0**	**4,3**	**P<0,01***
TMC (µm)	164,2	162,4	NS

2.1.2. Modification de la composition en acides gras

Le tableau 2 présente la teneur en acides gras des laits (%, g pour 100 g des acides gras totaux) pour lesquels l'effet du polymorphisme au locus *DGAT1* est significatif (seuil de significativité fixé à 10%). La teneur en C7 :0, C14 :1, C15 :0 et C16 :1 *cis*9 des individus de génotype « faible » est significativement inférieure à celle des individus de génotype « fort ». En revanche, avec le génotype « faible » la teneur est plus élevée en C14 :0, en C18 mono et poly-insaturés et en C20 mono et poly-insaturés.

Tableau 2 : Polymorphisme au locus *DGAT1* et composition en acides gras des laits

Acide gras (%)	AA/AA (K232)	GC/GC (A232)	P-value
C7:0	**0,05**	0,04	P<0,05
C14:0	12,91	**13,87**	P<0,05
C14:1	**1,20**	1,01	P<0,001
C15:0	**1,56**	1,40	P<0,1
C16:1 cis9	**1,55**	1,20	P<0,05
C18:1 cis11	0,47	**0,56**	P<0,1
C18:1 cis12	0,17	**0,22**	P<0,1
C18:1 cis16	0,03	**0,04**	P<0,1
C18:2 n-6	1,17	**1,41**	P<0,05
C18:3 n-3	0,28	**0,34**	P<0,05
C20:1 cis11	0,01	**0,02**	P<0,01
C20:2 n-6	0,02	**0,03**	P<0,1

2.1. EFFET DU POLYMORPHISME AU LOCUS *DGAT1* SUR L'ACTIVITE TRANSCRIPTIONNELLE DU TISSU MAMMAIRE

En fixant le seuil de significativité à 10%, 1 991 cibles sur 43 061 sont déclarées différemment exprimées entre les deux génotypes au locus *DGAT1*. Ces cibles correspondent à 1 496 gènes différents ayant une annotation acceptée par IPA qui se répartissent en 698 gènes sur-exprimés et 798 gènes sous-exprimés pour le génotype « faible » en TB. Les principales voies métaboliques affectées par le polymorphisme au locus *DGAT1* sont le « métabolisme des lipides » (124 gènes) et la « biochimie des petites molécules » (168 gènes). A partir de la synthèse de Bionaz et Loor (2008), une sélection des gènes les plus pertinents de ces deux voies a été effectuée et est présentée à la figure 1.

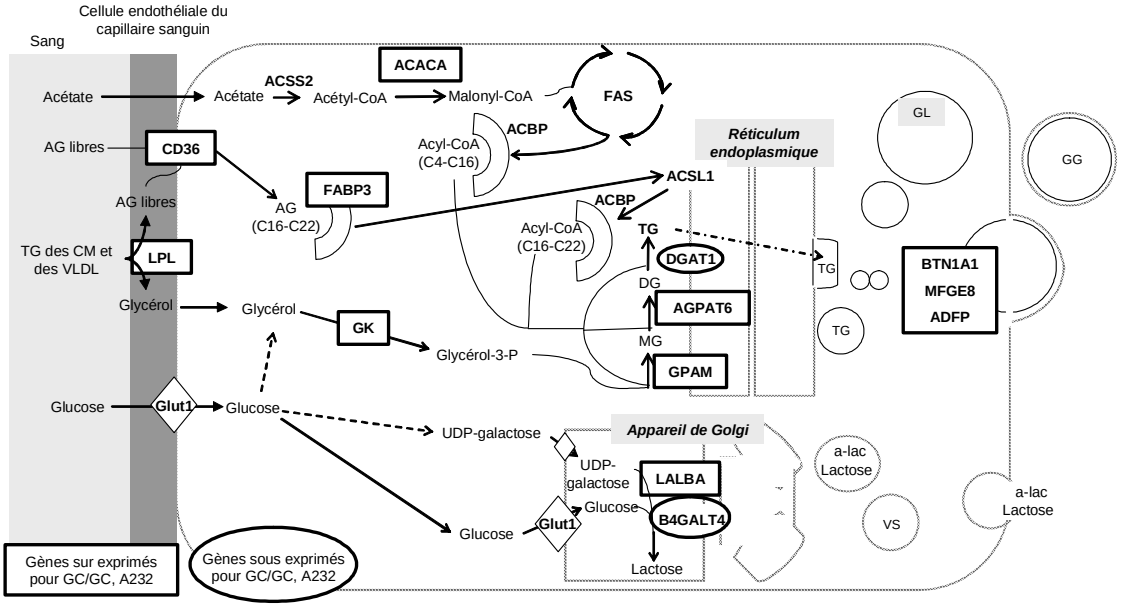


Figure 1 : Effet du polymorphisme au locus DGAT1 sur l'expression dans le tissu mammaire des gènes des voies de synthèse du lactose et des lipides du lait. (ACACA : acétyl-CoA carboxylase, ACBP : acyl-CoA binding protein, ACSL : l'acyl-CoA synthétase chaîne longue, ACSS : l'acyl-CoA synthétase chaîne courte, ADFP : adipophiline, AG : acides gras, AGPAT6 : acylglycerol-3-phosphate acyltransférase, BTN1A1 : butyrophilline, B4GALT4 : galactosyl transférase, CM : chylomicron, DG : diacylgycérol, DGAT : diacylglycerol acyltransferase, FABP : fatty acid binding protein, FAS : fatty acid synthétase, GG : globule gras, GK : glycérol kinase, GL : gouttelette lipidique, Glut1 : transporteur de glucose, GPAM : glycerol-3-phosphate acyltransférase, LALBA : alpha-lactalbumine, MFGE8 : milk fat globule-EGF factor 8, MG : monoacylglycérol, TG : triglycérides, VLDL : Very Low Density Lipoprotein, VS : vésicule de sécrétion, α-lac : alpha-lactalbumine).

Les gènes spécifiant les protéines de captation et de transport des acides gras longs telles que la lipoprotéine lipase (LPL), CD36, et la fatty acid binding protein (FABP3) sont sur-exprimés chez les individus de génotype « faible ». Il en va de même pour les gènes spécifiant l'acétyl CoA carboxylase (ACACA) qui intervient dans la synthèse *de novo* des acides gras, la glycerol-3-phosphate acyltransférase (GPAM) et l'acylglycerol-3-phosphate acyltransférase (AGPAT6) qui interviennent dans la synthèse du diacylglycérol. Le gène spécifiant l'α-lactalbumine (LALBA), la protéine nécessaire à la synthèse du lactose est également sur-exprimé chez les individus de génotype « faible » qui en revanche sous-expriment les gènes spécifiant DGAT1 et la glucosyltransférase (B4GALT4).

3. DISCUSSION

L'allèle DGAT1 « faible » (A232) est associé à une diminution des acides gras à chaînes courtes et moyennes (C4 à C12) et une augmentation des acides gras à chaînes longues et insaturés (>C18), ce qui va dans le sens des résultats publiés récemment par nos collègues néerlandais (Schennink et al., 2007). Les acides gras à chaîne carbonée supérieure à 16 atomes de carbone sont essentiellement d'origine alimentaire (ou viennent du tissu adipeux lors d'un bilan énergétique négatif) alors que les acides gras à chaîne courte et moyenne sont synthétisés *de novo* dans la cellule épithéliale mammaire (Chilliard et al., 2001). La sur-expression pour les animaux de génotype « faible » des gènes de la LPL, le CD36 et la FABP3 qui facilitent la captation et le transport des acides gras à chaîne longue (de C16 à C22) corrobore les résultats enregistrés en matière de composition en acides gras des laits (Bernard et al., 2008). La sous-expression du gène DGAT1 pour les individus de génotype « faible » est en revanche observée pour la première fois. Elle renforcerait l'effet de la mutation K232A rendue responsable d'une perte de l'efficacité de l'enzyme qui catalyse l'étape limitante de la synthèse des triglycérides favorisant ainsi l'accumulation de diacylglycérols, ce qui permettrait la synthèse accrue de phospholipides membranaires (Bell et al., 1981). La diminution de la taille des globules gras pour le génotype « faible » et donc l'augmentation de la surface membranaire secrétée avec le globule gras va dans le sens d'un telle hypothèse. Par ailleurs, l'absence d'effet significatif du génotype sur la matière grasse laitière s'explique par le fait que l'augmentation de la production laitière avec les individus de génotype « faible » compenserait l'effet significatif observé sur la baisse du taux butyreux. Cette augmentation est à rapprocher de la sur-expression du gène spécifiant la protéine nécessaire à la voie de synthèse du lactose (LALBA). Cette observation est en accord avec les travaux menés chez la souris sur l'invalidation du gène DGAT1 qui montrent que la perturbation de la synthèse des triglycérides s'accompagne d'une augmentation de la dépense énergétique et du métabolisme glucidique de l'adipocyte (Chen, 2006). Ces éléments suggèrent que le glucose non utilisé pour la synthèse des triglycérides du lait pourrait servir au métabolisme glucidique qui est principalement la voie de synthèse du lactose dans le tissu mammaire.

La validation des différentiels d'expression de certains gènes est en cours (PCRq) au laboratoire.

CONCLUSIONS

Notre étude a permis de confirmer que le polymorphisme affectant le locus *DGAT1* analysé ici, modifie la teneur du lait en matière grasse et de mettre en évidence qu'il perturbe le fonctionnement du tissu mammaire dans un sens qui pourrait ralentir la synthèse des triglycérides tout en favorisant celle du lactose. Il semble que la matière grasse des laits des animaux homozygotes pour l'allèle « faible » (GC, A232) contient plus d'acides gras d'origine alimentaire ou adipeuse (C18-C20). Ainsi, ces animaux pourraient répondre davantage aux variations de l'alimentation, ou bien mobiliser davantage leurs réserves corporelles. Ces hypothèses vont pouvoir être vérifiées prochainement à la faveur de l'expérimentation menée dans le cadre du programme « Genomilk Fat » qui combinera l'analyse des effets du génotype et du régime alimentaire.

Les travaux présentés ici font partie intégrante du programme « Genomilk Fat » co-financé par l'ANR (Genanimal) et Apis-Gene. Le projet a également reçu un support financier du Ministère de l'Agriculture (CAS DAR). Les auteurs souhaitent remercier le personnel des unités expérimentales INRA du Pin-au-Haras et de Theix pour le soin apporté aux animaux et l'organisation des divers prélèvements, Didier Goidin de la société Agilent Technologies pour les conseils techniques lors de l'utilisation des lames 4x44K bovines et le personnel du CRB GADIE de l'INRA de Jouy-en-Josas pour la production et la fourniture des lames oligoarrays 22K bovines. Nos remerciement s'adressent aussi à Jean-Claude Mocquot, Pierre-Louis Gastinel et Sophie Mattalia de l'Institut de l'Elevage pour l'initialisation et le suivi de la collaboration INRA-Institut de l'Elevage dans le domaine de la génomique fonctionnelle, à Joëlle Leonil, Jean-Louis Peyraud, Yves Chilliard et Didier Boichard de l'INRA pour leur participation active au programme « Genomilk Fat ».

Bell, R.M., Ballas, L.M., Coleman, R.A., 1981. J. Lipid Res. 22 :391-403

Benjamini, Y., Hochberg, Y., 1995. J. R. Statist. Soc. B, 57, 289-300

Bernard, L., Leroux, C., Chilliard, Y., 2008. Adv. Exp. Med. Biol., 606, 67-108

Bingham, S., 2002. Essentials of Human Nutrition., 335-54.

Bionaz, M., Loor, J.J., 2008. BMC Genomics, 9 :366

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, et al., 2003. Genet. Sel. Evol., 35, 77-101

Chen, H.C., 2006. Nutr. Metabolism, 3, 10

Chilliard, Y., Ferlay, A., Doreau, M., 2001. INRA Prod. Anim, 14, 323-335

Coleman, R.A., Lewin, T.M., Muoio, D.M., 2000. Annu. Rev. Nutr, 20, 77-103

Coppieters, W., Riquet, J., Arranz, J.J., Berzi, P., et al., 1998. Mamm Genome., 9, 540-544

Delmar, P., Robin, S., Daudin, J.J., 2005. Bioinformatics, 21, 502-508

Farr, V.C., Stelwagen, K., Cate, L.R., Molenaar, A.J., McFadden, T.B., Davis, S.R., 1996. J.Dairy Sci., 79, 543-549

Grisart, B., Coppieters, W., Farnir, F., Karim, L., et al., 2002. Genome Res., 12, 222-231

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, et al., 2004. Proc. Natl. Acad. Sci., 101, 2398-2403

Larroque, H., Gallard, Y., Boichard, D., Colleau, J.J., 2003. Renc. Rech. Ruminants, 10, 61

Schennink, A., Stoop, W.M., Visker, M.H.P.W., Heck, et al., 2007. Anim. Genet., 38, 467-473

Smith, S.J., Cases, S., Jensen, D.R., Chen, H.C., et al., 2000. Nat. Genet., 25, 87-90

Annexe 3 : Analyse des puces à ADN

Analyse des lames de verre après quantification avec le logiciel Feature Extraction (Agilent Technologies)

Ce document a pour objectif de rendre plus facile l'analyse des données issues du logiciel de quantification Feature Extraction développé et commercialisé par Agilent Technologies. La partie analyse différentielle répond seulement à la question comparaison du transcriptome d'une condition A avec le transcriptome d'une condition B. Pour toute autre question ou plan expérimental, il faut consulter un biostatisticien.

1. Installation de R et Tinn-R sur un ordinateur

1.1. Tinn-R

Logiciel de programmation. Les programmes s'enregistrent sous la forme « fichier.r ». C'est dans Tinn-R que l'on peut créer, ouvrir, modifier, enregistrer et lancer des programmes. C'est un éditeur de code en langage S gratuit.
Le téléchargement de Tinn-R peut se faire à partir du lien suivant : http://www.sciviews.org/Tinn-R/

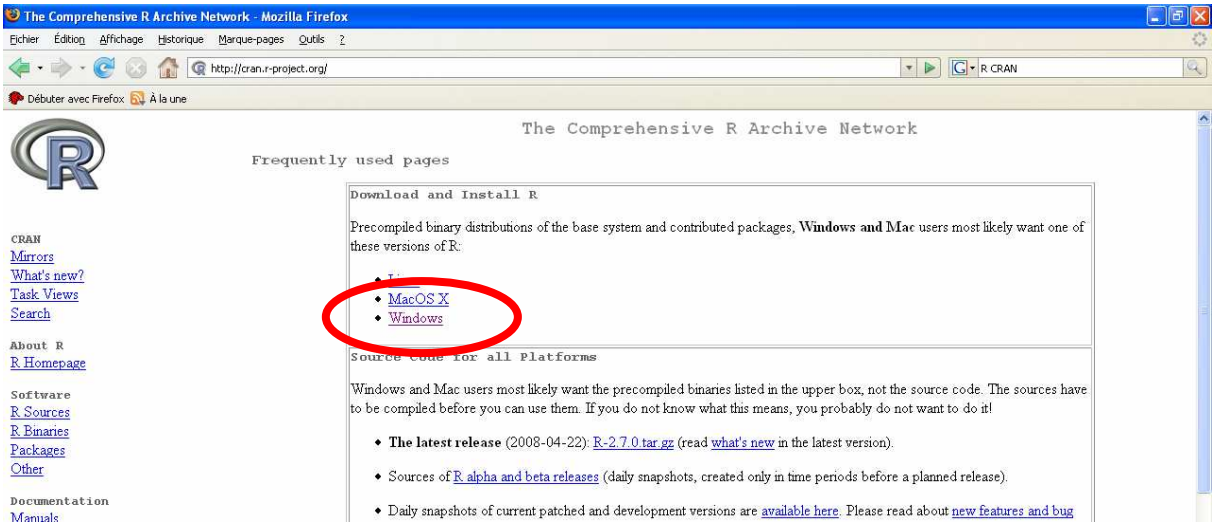
Pour Windows, la dernière version à télécharger se trouve sur le lien suivant :
https://sourceforge.net/projects/tinn-r

1.2. R

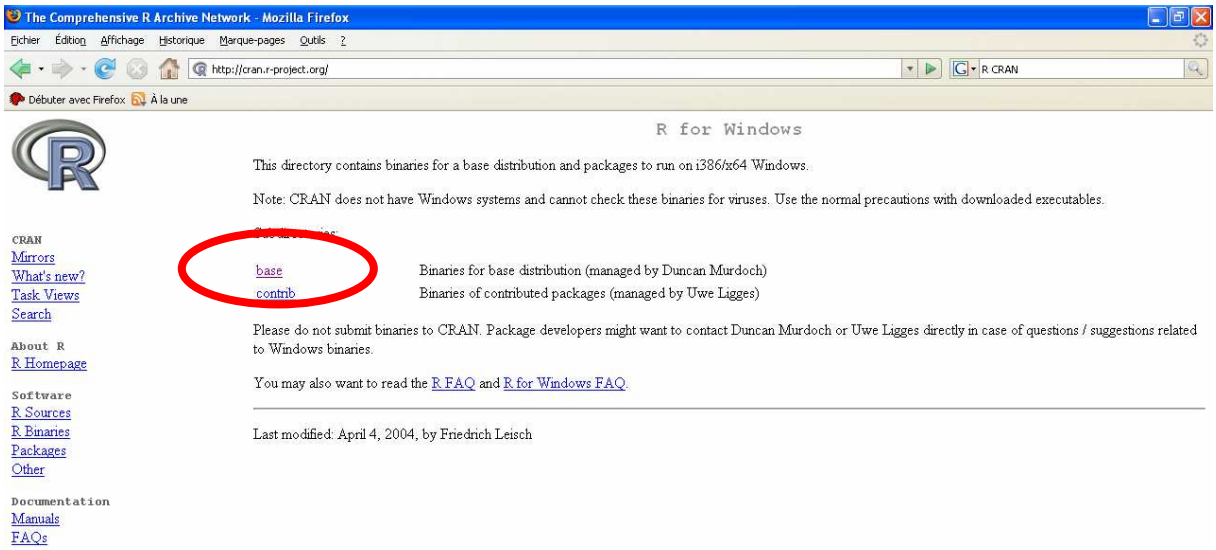
Logiciel de calcul libre utilisant le langage S. Sert beaucoup pour les analyses statistiques et graphiques.
Les fichiers pour installer R sont disponibles sur le site Comprehensive R Archive Network (CRAN).
http://cran.r-project.org/

Pour installer R sous Windows, suivez les instructions suivantes :

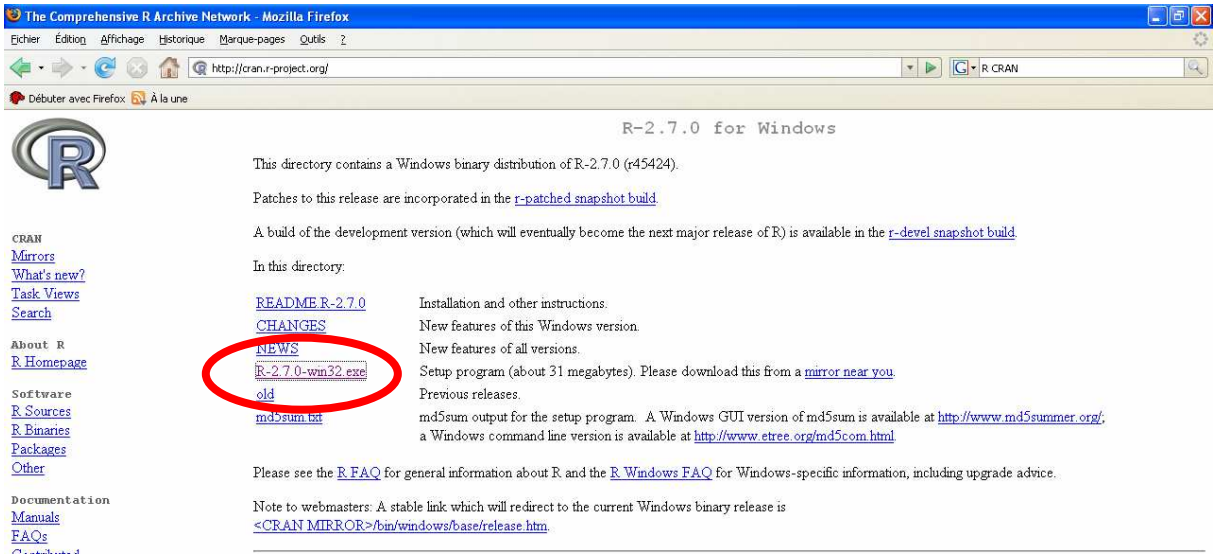
- Sous le CRAN, cliquer sur Windows



- Puis cliquer sur « Base » (installation des packages de base de R)



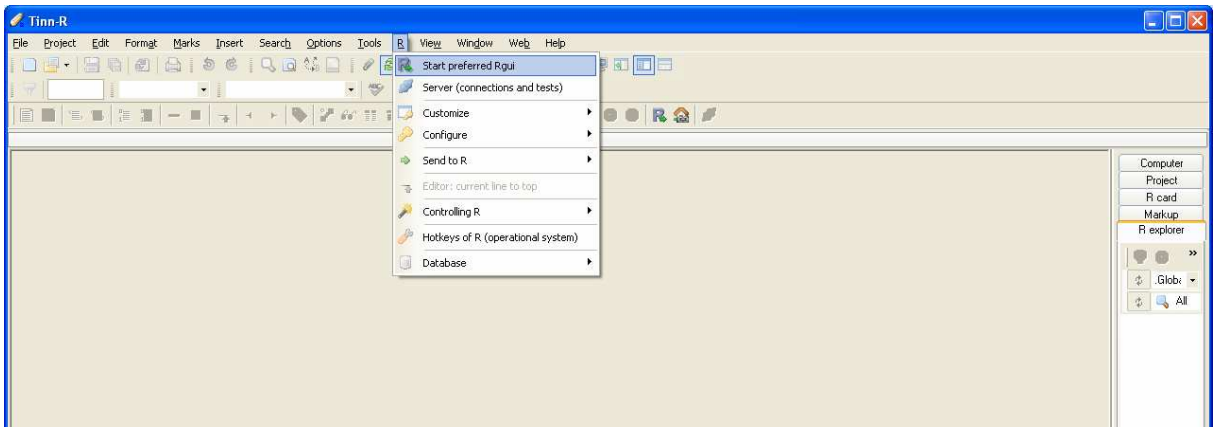
- Puis cliquer sur le « .exe » qui permet de télécharger R sur votre ordinateur. Suivez les instructions.



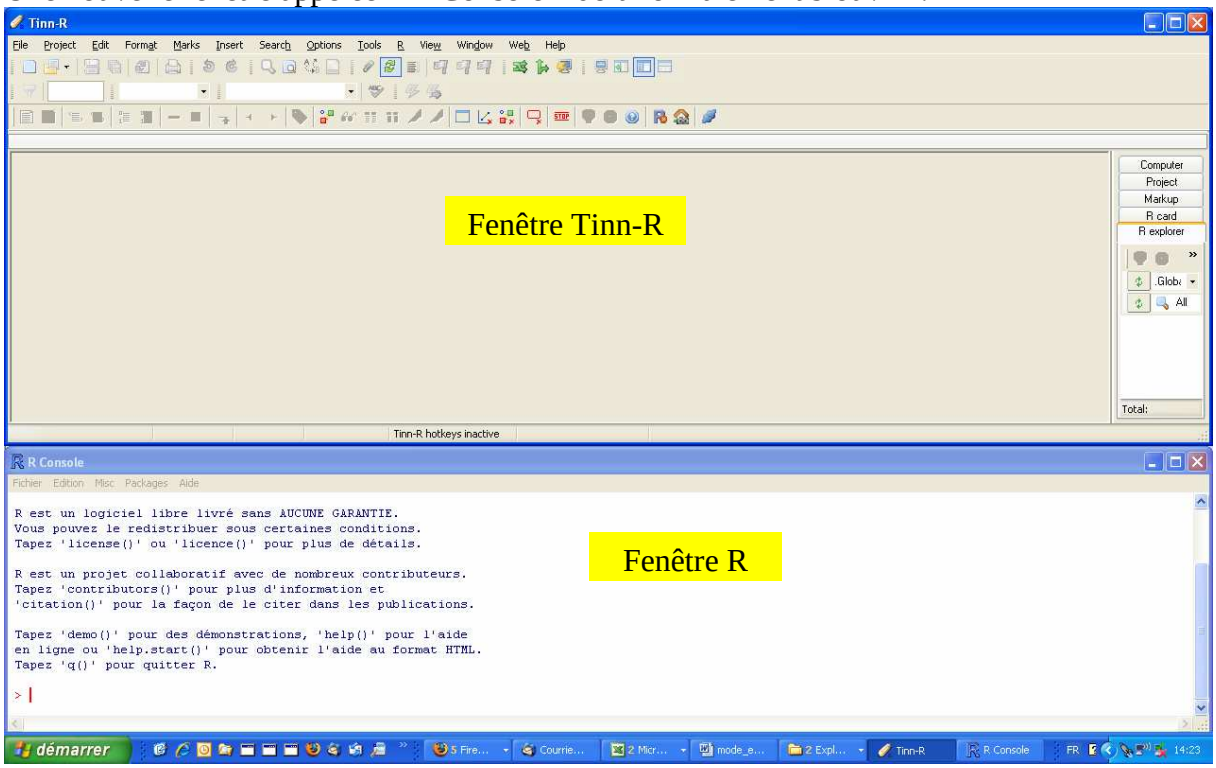
Pour plus de renseignements sur le fonctionnement direct de R, ne pas hésiter à interroger Google en tapant « R pour les débutants ». De très bons documents francophones sont alors accessibles et très compréhensifs.

Pour faire fonctionner R et Tinn-R, il faut d'abord ouvrir Tinn-R (soit en cliquant sur l'icône sur votre bureau, soit en ouvrant un fichier du type *.R, *.r, *.Q or *.q) puis ouvrir R à partir de Tinn-R en suivant les instructions suivantes :

- cliquer sur R
- puis cliquer sur « start preferred Rgui »



Une nouvelle fenêtre appelée « R Console » doit normalement s'ouvrir :



2. Normalisation

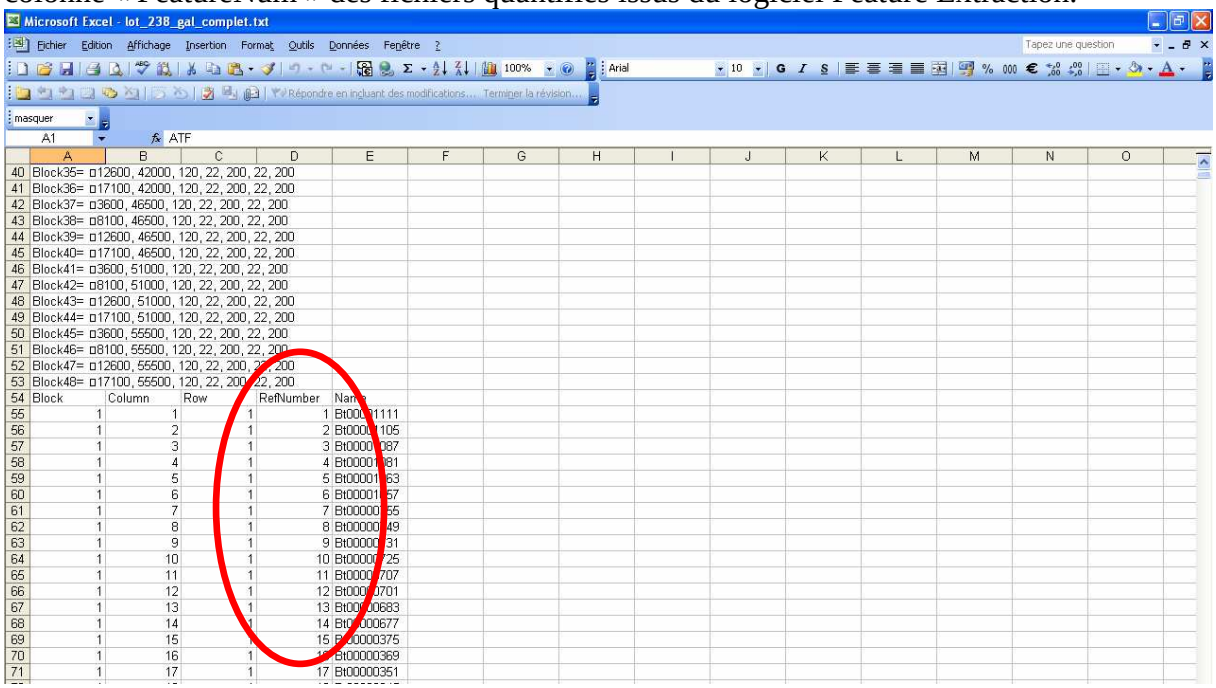
La normalisation des données dans ce cas consiste en une normalisation loess générale (Yang et al., 2002) suivi d'une correction de l'effet bloc par une soustraction de la valeur médiane d'un bloc lorsqu'il y a une colonne bloc (fichier des lames type CRB).

2.1.Lames de type CRB (avec un point gal)

2.1.1. Préparation du fichier «.gal » pour identifier les spots

Souvent le fichier gal utilisé pour la quantification dans Feature Extraction est très épuré. Pendant la normalisation il est possible de rajouter des informations sur les lignes qui se trouveront dans le fichier issu de la normalisation. Pour cela il faut compléter le fichier gal qui sert à la normalisation.

En plus des colonnes qui nous intéressent, il faut rajouter une colonne « RefNumber » comportant des chiffres de 1 au nombre de spots de la lame. Cette colonne est comparable à la colonne « FeatureNum » des fichiers quantifiés issus du logiciel Feature Extraction.

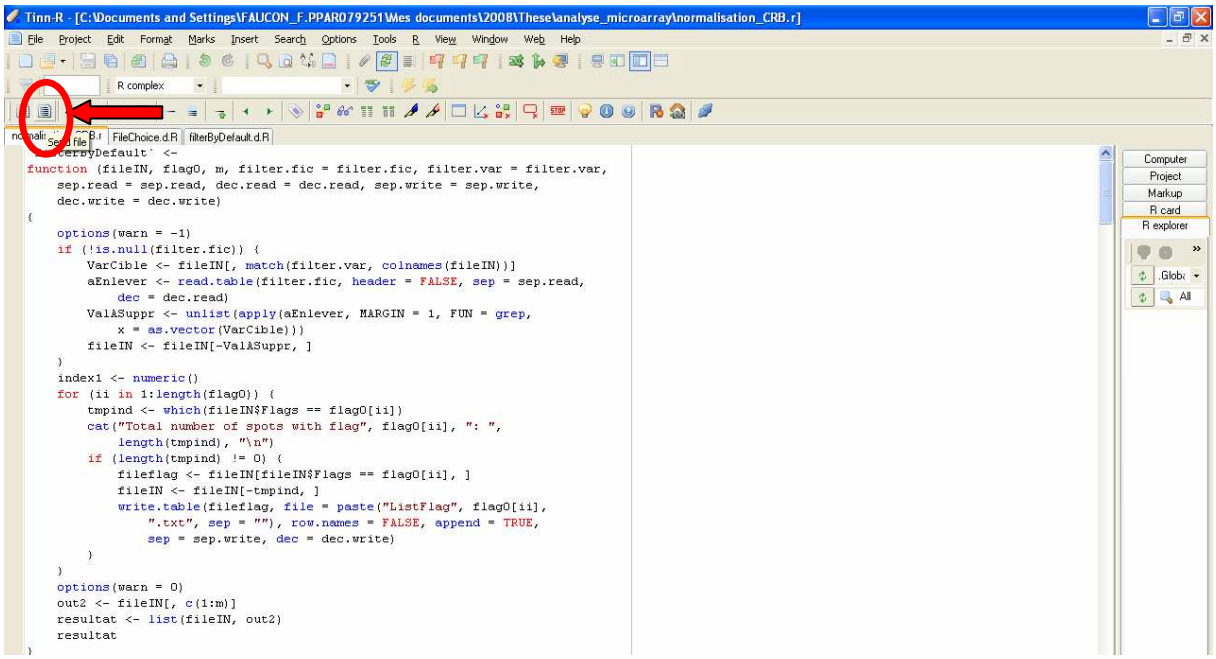


ATTENTION : penser à enregistrer ce fichier gal dans le même dossier que les fichiers quantifiés qui vont servir pour la normalisation.

2.1.2. Utilisation de la fonction normalisation

Pour la normalisation des lames type CRB (avec fichier gal), il faut ouvrir le fichier normalisation_CRB.r dans Tinn-R.
Ce fichier a été crée à partir des fonctions filterByDefault, FileChoice et normalisation du package anapuce développé et publié dans le Bioconductor (http://www.bioconductor.org/) par Julie Aubert (AgroParisTech). La fonction normalisation a alors été adapté par Julie Aubert aux lames du type celles du CRB.

- lancer la fonction en cliquant sur l'icône suivante :



- Puis dans la fenêtre de R Console (c'est celle du bas !) taper la commande normalisation.CRB(ATTENTION, ici ne changer que les options que vous avez besoin de changer, la liste des options ainsi que les valeurs par défaut sont ci dessous)
  Liste des options de la fonction normalisation.CRB et valeurs par défaut

fileIN : fichier quantifié à analyser. Par défaut = NULL. Nous conseillons de ne rien changer pour cette option car la fonction normalisation.CRB permet la recherche des fichiers à normaliser.
var1 : signal rouge à prendre dans le fichier quantifié. Par défaut « rMedianSignal »
var2 : signal vert à prendre dans le fichier quantifié. Par défaut « gMedianSignal »
n : nombre de colonnes qui décrivent un spot. Par défaut « 5 »
flag : types de flag à prendre en compte. Par défaut « c(1 :8) ». Pour la signification des numéros de flag, se référer à la remarque ci dessous.
graph : si vous voulez que la fonction génère des graphes. Par défaut « TRUE »
sep.write : séparateur des tableaux générés. Par défaut « \t »
sep.read : séparateur des tableaux lus. Par défaut « \t »
dec.write : décimale des tableaux générés. Par défaut « . »
dec.read : décimale des tableaux lus. Par défaut « . »
filter.function : fonction à appliquer avant de commencer la normalisation. Par défaut « filterByDefaut »
filter.fic : argument de la fonction filter.function. Par défaut « NULL »
filter.var : argument de la fonction filter.function. Par défaut « NULL »
file.gal : nom du fichier du point gal préparé comme décrit au 2.1.1 de ce document. A CHANGER IMPERATIVEMENT !!!

- Cliquer sur entrer
- Une fois la fonction lancée, une interface vous demande où se trouvent les fichiers quantifiés. Faire une recherche pour préciser. Attention à ce que le fichier gal se trouve au même endroit que les fichiers quantifiés.
  Ensuite il y a interrogation des fichiers que l'on souhaite normaliser, mettre « y » (yes) lorsque l'on veut normaliser le fichier, « n » (no) lorsque l'on ne veut pas.
- Une question est ensuite posée concernant l'ordre des fichiers qu'il faut analyser. ATTENTION, l'ordre dans lequel les fichiers sont rentrés a toute son importance. Penser à mettre les deux lames d'un même dye-swap l'une après l'autre. Dans le cas de plusieurs

répétitions biologiques ou d'un plan en référence, pensez toujours à mettre la lame portant la condition A (ou la référence) en rouge en premier et la lame portant la condition B rouge (ou les conditions comparés à la référence) en deuxième. L'ordre a vraiment beaucoup d'importance, cette étape nécessite beaucoup de temps et d'avoir le plan d'hybridation à porté de main.

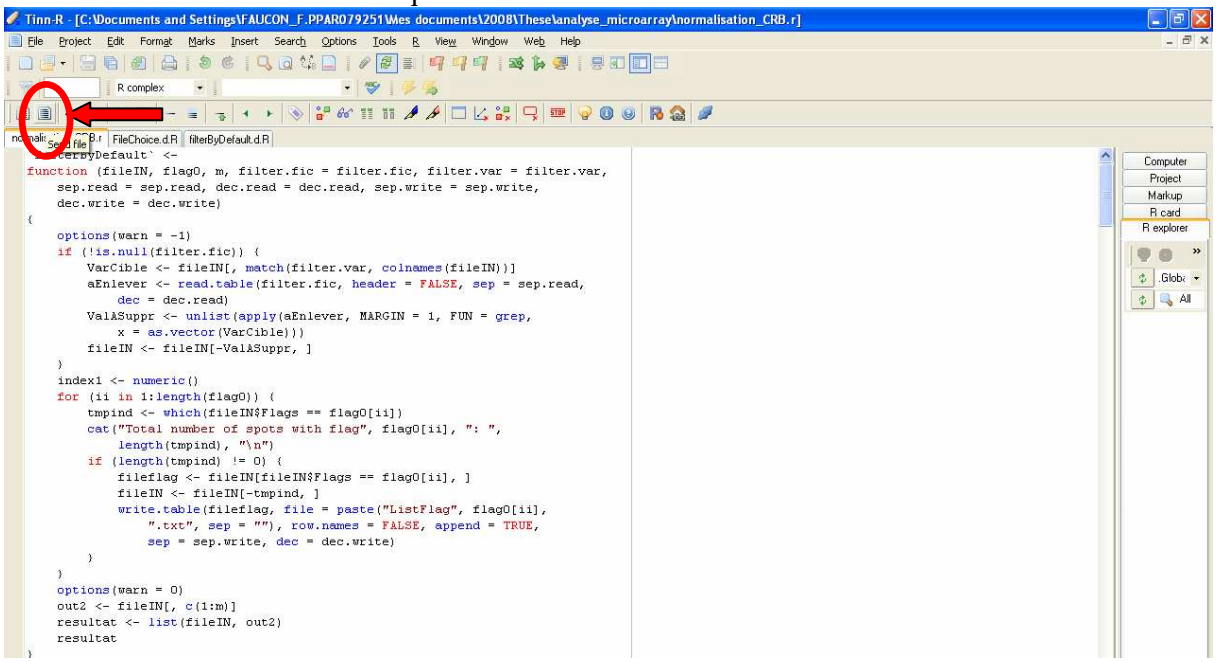
Remarque : signification des numéros de Flags.

1* gIsFeatNonUnifOL 2* rIsFeatNonUnifOL	Feature uniforme ou pas. Il est non uniforme quand le bruit des pixels excède un seuil établit pour un feature uniforme
3* gIsFeatPopOL 4* rIsFeatPopOL	Population outlier ou pas. Lorsque son signal est inférieur à un seuil ou excède un seuil déterminés en utilisant une fois l'interquartile
5* gIsBgNonUnifOL 6* rIsBgNonUnifOL	Même chose mais pour le bruit de fond
7* gIsBGPopnOL 8* rIsBGPopnOL	Même chose mais pour le bruit de fond

2.2. Lames du type Agilent

Pour la normalisation des lames type lame Agilent, il faut ouvrir le fichier normalisation_Agilent.r dans Tinn-R.
Ce fichier a été crée à partir des fonctions filterByDefault, FileChoice et normalisation du package anapuce développé et publié dans le Bioconductor (http://www.bioconductor.org/) par Julie Aubert (AgroParisTech). La fonction normalisation a alors été adapté par Julie Aubert aux lames du type celles d'Agilent.

- lancer la fonction en cliquant sur l'icône suivante :



- Puis dans la fenêtre de R Console (c'est celle du bas !) taper la commande normalisation.Agilent(ATTENTION, ici ne changer que les options que vous avez besoin de changer, la liste des options ainsi que les valeurs par défaut sont ci dessous).
  Liste des options de la fonction normalisation.Agilent et valeurs par défaut

fileIN : fichier quantifié à analyser. Par défaut = NULL. Nous conseillons de ne rien changer pour cette option car la fonction normalisation.Agilent permet la recherche des fichiers à normaliser.

Red : signal rouge à prendre dans le fichier quantifié. Par défaut « rMedianSignal »
Green : signal vert à prendre dans le fichier quantifié. Par défaut « gMedianSignal »
n : nombre de colonnes qui décrivent un spot. Par défaut « 14 »
flag : types de flag à prendre en compte. Par défaut « c(1 :8) ». Pour signification des numéros, cf. remarque ci dessus.
graph : si vous voulez que la fonction génère des graphes. Par défaut « TRUE »
sep.write : séparateur des tableaux générés. Par défaut « \t »
sep.read : séparateur des tableaux lus. Par défaut « \t »
dec.write : décimale des tableaux générés. Par défaut « . »
dec.read : décimale des tableaux lus. Par défaut « . »
filter.function : fonction à appliquer avant de commencer la normalisation. Par défaut « filterByDefaut »
filter.fic : argument de la fonction filter.function. Par défaut « NULL »
filter.var : argument de la fonction filter.function. Par défaut « NULL »
quote.read : détermine les délimitations de texte dans les fichiers lus. Par défaut «"\"" »

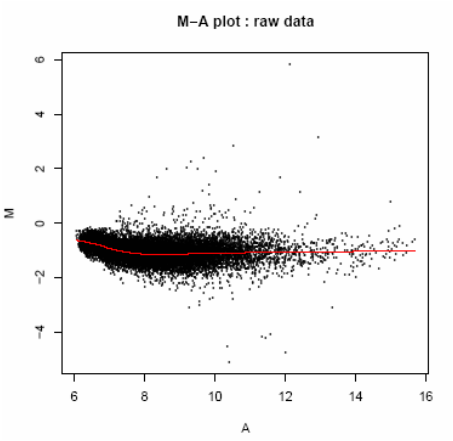
- Cliquer sur entrer
- Une fois la fonction lancée, une interface vous demande où se trouve les fichiers quantifiés. Faire une recherche pour préciser.
  Ensuite il y a interrogation des fichiers que l'on souhaite normaliser, mettre « y » (yes) lorsque l'on veut normaliser le fichier, « n » (no) lorsque l'on ne veut pas.

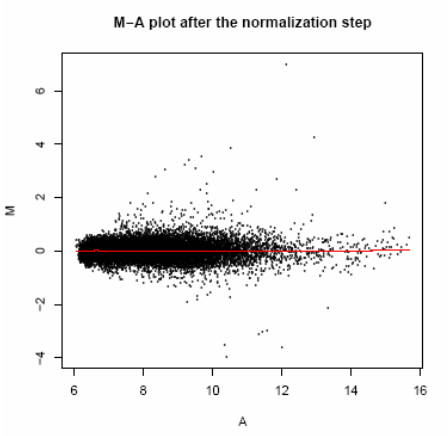
- Une question est ensuite posée concernant l'ordre des fichiers qu'il faut analyser. ATTENTION, l'ordre dans lequel les fichiers sont rentrés a toute son importance. Penser à mettre les deux lames d'un même dye-swap l'une après l'autre. Dans le cas de plusieurs répétitions biologiques ou d'un plan en référence, pensez toujours à mettre la lame portant la condition A (ou la référence) en rouge en premier et la lame portant la condition B rouge (ou les conditions comparés à la référence) en deuxième. L'ordre a vraiment beaucoup d'importance, cette étape nécessite beaucoup de temps et d'avoir le plan d'hybridation à porté de main.

2.3. Fichiers générés par les fonctions normalisation

Les fonctions normalisation génèrent :
- Un fichier txt dans lequel se trouve plusieurs colonnes.
Les n premières colonnes correspondent au nombre de colonnes descriptives du spot (à rentrer dans les options au moment de l'application de la fonction).
Ensuite pour chaque lame i il y a 4 colonnes qui se succèdent :
Ai : log ratio moyen ou intensités moyennes A = (log2[Cyanine 5]+ log[Cyanine 3]) / 2
M.normi : log du ratio rouge/vert M = log2(Cyanine 5) – log(Cyanine 3)
corLowessi : correction loess
corBlocki : correction par bloc (non existant pour les fichiers issus de normalisation.Agilent)

- Les M-A plots pour chaque lame représentant en abscisse la valeur A, en ordonnées la valeur M.norm. La courbe Loess est aussi représentée. Elle doit être la plus linéaire possible après la normalisation.





3. Vérification de la cohérence du jeu de données

3.1. Verifdonnee

La vérification du jeu de données se fait en appliquant la fonction verifdonnee au fichier resNorm.txt issu des normalisations décrites si dessus. L'objectif de cette fonction est de vérifier de façon rapide qu'il n'y a pas de grosses erreurs dans le jeu de données (lames rentrées dans le désordre, swaps qui n'ont pas fonctionnés…) qui risqueraient d'affecter les résultats des futures analyses. Cette fonction a été développée entièrement sous R par Félicie FAUCON en début 2008.

Le conseil est de copier le fichier resNorm.txt dans un dossier différent de celui de la normalisation pour faire la vérification du jeu de données.

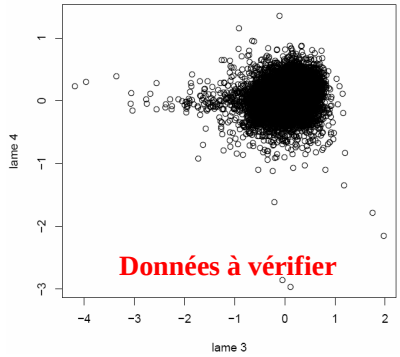
Pour appliquer la fonction verifdonnee, la procédure est simple.

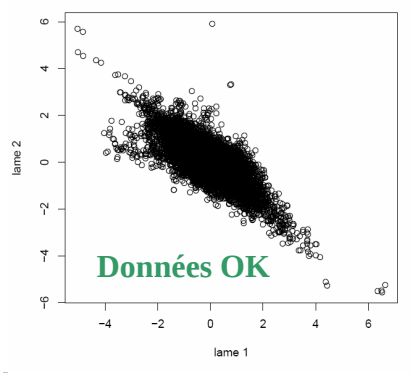
- ouvrir le fichier verifdonnee.r dans Tinn-R.
- après avoir lancé R, lancer la fonction verifdonnee en cliquant sur l'icône de Tinn-R.
- se placer dans la fenêtre R Console cliquer sur Fichier, puis sur Changer le répertoire courant… et se placer dans le dossier où se trouve le fichier resNorm.txt à vérifier.
- Puis taper la commande verifdonnee(options à modifier)

Les options possibles pour cette fonctions sont :
FileIN : fichier de données à vérifier. Par défaut « resNorm.txt »
var1 : log ratio. Par défaut « M.norm »
var2 : moyenne des log. Par défaut « A »
graph : si on veut que des graphes soient générés. Par défaut « TRUE »
sep.read : séparateur des fichiers lus. Par défaut « \t »
n : nombre de colonnes descriptives du spot. Par défaut « 5 » (adapté à un fichier type lame CRB)
nb.col.lame : nombre de colonnes présentes dans le fichier resNorm.txt corespondant à 1 lame. Par défaut « 4 ». C'est adapté à un fichier type lame CRB. Pour le fichier type lame Agilent il faut mettre 3 (car pas de colonne corBlock pour les lames).
sep.write : séparateur de création de tableau. Par défaut « \t »

Les fichiers générés par cette fonction sont :

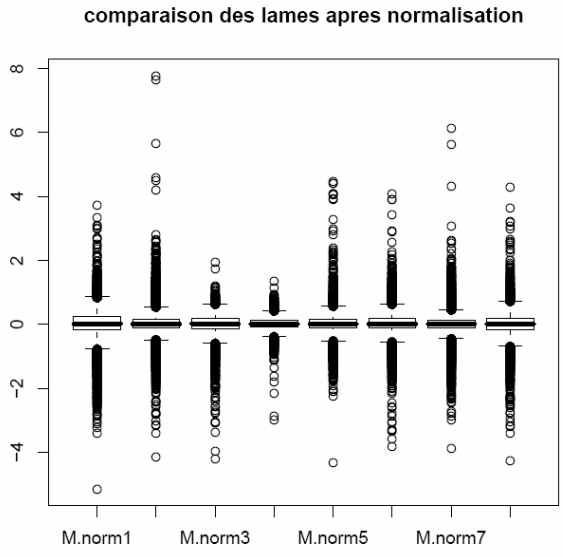
- 1 fichier pdf par dye-swap « graphcorlamei » représentant la répartition des log ratio entre deux lames d'un même dye-swap. Le mieux est que cette répartition soit la plus linéaire possible et descendante.





- un fichier pdf « complame » qui compare toutes les lames du fichier resNorm.txt. Il faut vérifier que toutes les lames soient bien centrées sur zéro et que deux lames successives (les deux lames d'un même dye-swap) aient à peu près le même profil.

Centrage des lames sur zéro ⇨



- un fichier « corlamenorm.txt » contenant les coefficients de corrélation (r) de chaque lame avec toutes les autres lames du projet. Il vaut mieux que le coefficient de corrélation entre deux lames d'un même dye-swap soit le plus proche possible de -1 (les lames sont alors bien reversées). Lorsque ceci n'est pas le cas (lame 3 et 4 dans le tableau suivant), cela peut venir de différentes choses à vérifier :

les lames n'ont pas été quantifiées correctement,
les lames n'ont pas été rentrées dans le bon ordre au moment de la normalisation,
un des marquages sur l'une des lames n'a pas fonctionné,
les deux conditions hybridées sur ces lames sont très proches, dans ce cas les points du graphe « graphcorlamei » doivent être centrés sur zéro,
bien d'autres explications à chercher dans le jeu de données.

	M.norm1	M.norm2	M.norm3	M.norm4	M.norm5	M.norm6	M.norm7	M.norm8
M.norm1	1,000	-0,699	0,329	-0,069	0,663	-0,529	0,381	-0,318
M.norm2	-0,699	1,000	-0,096	0,408	-0,476	0,675	-0,205	0,487
M.norm3	0,329	-0,096	1,000	-0,164	0,315	-0,150	0,444	-0,012
M.norm4	-0,069	0,408	-0,164	1,000	-0,144	0,360	0,109	0,329
M.norm5	0,663	-0,476	0,315	-0,144	1,000	-0,709	0,253	-0,158
M.norm6	-0,529	0,675	-0,150	0,360	-0,709	1,000	-0,074	0,326
M.norm7	0,381	-0,205	0,444	0,109	0,253	-0,074	1,000	-0,625
M.norm8	-0,318	0,487	-0,012	0,329	-0,158	0,326	-0,625	1,000

- un fichier « corsignal.txt » contenant les coefficients de corrélation (r) de chaque signal avec tous les autres signaux. Pour calculer ces coefficients, la fonction crée un tableau (fichier « intsignal.txt ») qui contient les signaux rouge et vert normalisés pour chaque lame.

Dans une manipulation de puce à ADN, il y a 3 effets majoritaires : l'effet condition que l'on cherche à évaluer par la manipulation, l'effet couleur que l'on corrige en effectuant un dye swap et l'effet lame qu'il est difficile d'évaluer.

Le fichier « corsignal.txt » permet d'expliquer certaines choses dans le jeu de données mais les explications dépendront grandement du plan d'hybridation choisi.

Dans le cas d'un plan d'hybridation avec dye-swap, il est possible de vérifier que l'effet fluorochrome ne masque pas l'effet condition. Pour cela, il faut vérifier que le signal rouge d'une lame ait bien le plus fort coefficient de corrélation avec le signal vert de l'autre lame du même dye-swap (dans un dye-swap, la condition 1 est rouge sur la première lame du dye-swap et verte sur la deuxième lame du dye-swap, vu que logiquement le dye-swap correspond à l'inversion des fluorochromes).

	IntRed.norm1	IntGreen.norm1	IntRed.norm2	IntGreen.norm2	IntRed.norm3	IntGreen.norm3	IntRed.norm4	IntGreen.norm4
IntRed.norm1	1	0,966	0,961	0,986	0,964	0,954	0,938	0,945
IntGreen.norm1	0,966	1	0,986	0,969	0,98	0,983	0,972	0,976
IntRed.norm2	0,961	0,986	1	0,973	0,988	0,984	0,981	0,978
IntGreen.norm2	0,986	0,969	0,973	1	0,972	0,964	0,949	0,958
IntRed.norm3	0,964	0,98	0,988	0,972	1	0,989	0,983	0,987
IntGreen.norm3	0,954	0,983	0,984	0,964	0,989	1	0,983	0,984
IntRed.norm4	0,938	0,972	0,981	0,949	0,983	0,983	1	0,991
IntGreen.norm4	0,945	0,976	0,978	0,958	0,987	0,984	0,991	1

Dans cet exemple, il est évident qu'il y a un problème pour les lames 3 et 4.

Dans le cas d'un plan expérimental avec référence (même condition hybridées sur plusieurs lames), il est possible d'évaluer l'effet lame. En effet, la référence dans une couleur donnée sur une lame doit avoir un coefficient de corrélation élevé avec les autres lames où elle se trouve dans la même couleur.

3.2. Recherche des identifiants de points « bizarres » : selecpoint

Cette fonction permet de sélectionner sur un graphes de point des points « bizarres » puis de récupérer la liste de ces points.

Elle a été écrite en se servant des données issues de la normalisation du package anapuce (fichier resNorm.txt) en utilisant le graphe comparant les log-ratio des 2 lames d'un même dye-swap. Mais cette fonction peut s'adapter à d'autres types de données, il suffit d'adapter les options. Cette fonction a été développée entièrement sous R par Félicie FAUCON en début 2008.

Le conseil est de copier le fichier resNorm.txt ou le fichier dont vous voulez tracer les graphes et sélectionner les points dans un dossier différent de celui de la normalisation et de la vérification du jeu de données.

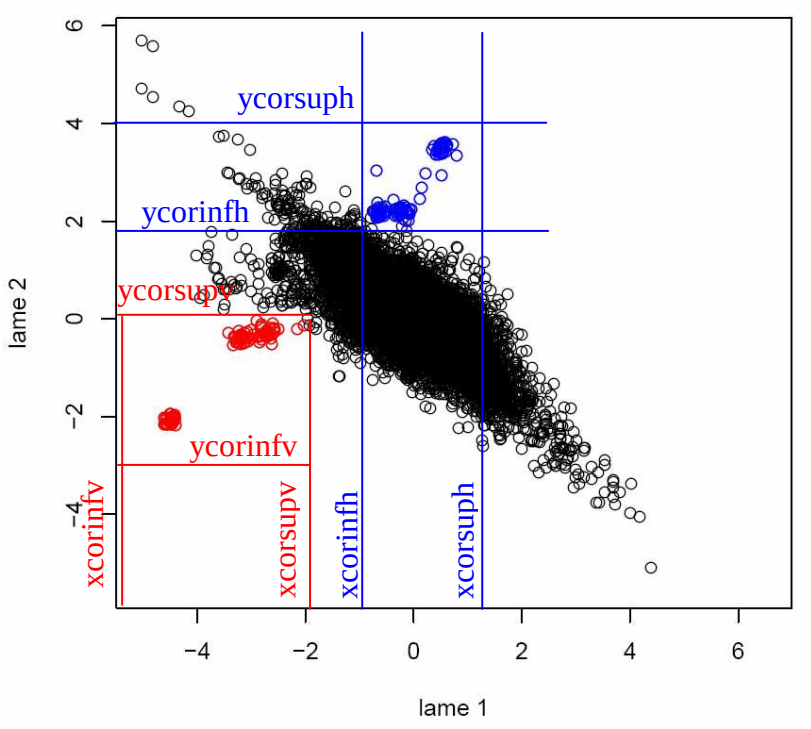
Pour appliquer la fonction selectpoint, la procédure est :

- ouvrir le fichier selectpoint.r dans Tinn-R.
- après avoir lancé R, lancer la fonction selectpoint en cliquant sur l'icône  de Tinn-R.
- se placer dans la fenêtre R Console cliquer sur Fichier, puis sur Changer le répertoire courant… et se placer dans le dossier où se trouve le fichier que vous voulez traiter.
- Puis taper la commande selectpoint(options à modifier)

Les options possibles pour cette fonctions sont :
FileIN : fichier de données contenant les variables avec lesquelles les graphes vont être tracés. Par défaut « resNorm.txt »
var1 : variable avec laquelle les graphes vont être tracés. Par défaut « M.norm »
var2 : pas utile a priori. Par défaut « A »
graph : si on veut que des graphes soient générés. Par défaut « TRUE »
sep.read : séparateur des fichiers lus. Par défaut « \t »
n : nombre de colonnes descriptives du spots. Par défaut « 5 » (adapté à un fichier type lame CRB)
nb.col.lame : nombre de colonnes présentes dans le fichier resNorm.txt corespondant à 1 lame. Par défaut « 4 ». C'est adapté à un fichier type lame CRB. Pour le fichier type lame Agilent il faut mettre 3 (car pas de colonne corBlock pour les lames).
sep.write : séparateur de création de tableau. Par défaut « \t »

xcorinfh
ycorinfh
xcorsuph
ycorsuph
ycorsuph
xcorinfv
ycorinfv
xcorsupv
ycorsupv

} Chiffres à rentrer selon le graphe ci-dessous



Les fichiers générés par cette fonction sont pour chaque dye-swap i :
- un fichier « spotverticaux i.txt » qui donne la liste des spots colorés en rouge sur le graphe
- un fichier « spothorizontaux i.txt » qui donne la liste des spots colorés en bleu sur le graphe
- un fichier « spotaberrant i.txt » qui donne la concaténation des deux listes précédentes
- un fichier « graphcorpointlamei.pdf » qui donne le graphe des corrélations avec en couleur les spots correspondants au deux listes précédentes.

4. Analyse différentielle (fonction DiffAnalysis du package anapuce)

L'analyse différentielle s'applique pour obtenir la liste des gènes différemment exprimés entre deux conditions. Sur la plateforme PICT, elle s'effectue à partir du fichier resNorm.txt issu de la normalisation des données en utilisant la fonction DiffAnalysis du package anapuce développé sous R par J. Aubert.

4.1. Téléchargement du package anapuce

Après avoir ouvert une fenêtre R Console, taper les commandes suivantes :
`source("http://bioconductor.org/biocLite.R")` %Permet de se connecter au serveur
`biocLite("anapuce")` %Téléchargement du package anapuce

4.2. La fonction DiffAnalysis

Cette fonction permet d'effectuer une analyse différentielle sur des données normalisées et appareillées (lame 2 couleurs) en utilisant différentes modélisation de la variance. Le résultat est des listes de gènes différemment exprimés entre deux conditions et calculés selon différents tests de comparaison multiple.

Pour appliquer la fonction DiffAnalysis :

- enregistrer le fichier resNorm.txt dans un nouveau dossier
- se mettre dans le fenêtre R Console
- puis changer le répertoire de travail en cliquant sur Fichier, puis sur Changer le répertoire courant… et se placer dans le dossier où se trouve le fichier resNorm.txt à traiter
- taper la commande library(anapuce) pour charger le package anapuce
- taper la commande DiffAnalysis(options à modifier)

Les options disponibles pour cette fonction sont :
fileIN : fichier de données à analyser. Par défaut « resNorm.txt »
n : nombre de colonnes descriptives du spot. Par défaut « 3 »
ind.array : les indices des lames à analyser. Par défaut « 1 :2 »
name.A : attention se mettre en caractère (c'est à dire entre guillemets). Nom de la colonne correspondant à la valeur de A (log ratio moyen A = (log[Cyanine 5]+ log[Cyanine 3]) / 2). Par défaut «« A »»
name.M : attention se mettre en caractère (c'est à dire entre guillemets). Nom de la colonne correspondant à la valeur de M (log du ratio rouge/vert M = log(Cyanine 5) – log(Cyanine 3)). Par défaut «« M.norm »».
fileOUT : nom du fichier de sortie des données. Par défaut « ListOfGenes.txt »
fileDelete : nom du fichier des gènes retirés de l'analyse. Par défaut « GenesOutOfAnalysis.txt »
procs : procédure d'ajustement des p-values à utiliser. Par défaut « c(« bonferroni » , « BH ») »
alpha : valeur des seuils de significativité à utiliser dans les procédures. Par défaut « c(0.05,0.05) »
dyeswap : indique si le design est en swap. Par défaut « TRUE »
indDS : index des lames correspondant aux swaps. Par défaut « c(2) »
function.trt : fonction à appliquer avant l'analyse des données. Par défaut « FALSE »
by.var : argument pour la fonction de l'option function.trt. Par défaut « ID »
varmixt.meth : argument logique indiquant si il faut utiliser le modèle de variance proposé par Delmar et al., 2005. Par défaut « FALSE »
header : indique si les fichiers contiennent le titre en première ligne. Par défaut « TRUE »
sep : le séparateur de champs à utiliser dans la fonction read.table. Par défaut « \t »
sep.write : le séparateur de champs à utiliser dans la fonction write.table. Par défaut « \t »
dec.write : signe à utiliser pour le point décimal. Par défaut « . »

Les fichiers générés par cette fonction sont :

- BH-*seuil*-ListOfGenes.txt : liste des gènes déclarés différemment exprimés avec une variance commune pour tous les gènes, modèle de correction des faux-positifs de Benjamini-Hochberg
- BH-*seuil*-VM-ListOfGenes.txt : liste des gènes déclarés différemment exprimés avec les gènes classés en groupe de même variance (Delmar et al., 2005), modèle de correction des faux-positifs de Benjamini-Hochberg
- bonferroni-*seuil*-ListOfGenes.txt : liste des gènes déclarés différemment exprimés avec une variance commune pour tous les gènes, modèle de correction des faux-positifs de Bonferroni
- bonferroni-*seuil*-VM-ListOfGenes.txt : : liste des gènes déclarés différemment exprimés avec les gènes classés en groupe de même variance (Delmar et al., 2005), modèle de correction des faux-positifs de Bonferroni
- CompleteListOfGenes.txt : liste complète des gènes analysés en prenant une variance commune pour tous les gènes

- GenesOutOfAnalysis.txt : liste des gènes retirés de l'analyse avec une variance commune pour tous les gènes (car variance trop différente)
- VarianceByGene-BH-*seuil*-ListOfGenes.txt : liste des gènes déclarés différemment exprimés avec une variance par gène, modèle de correction des faux-positifs de Benjamini-Hochberg
- VarianceByGene-bonferroni-*seuil*-ListOfGenes.txt : liste des gènes déclarés différemment exprimés avec une variance par gène, modèle de correction des faux-positifs de Bonferroni
- VM-CompeteListOfGenes.txt : liste complète des gènes analysés en regroupant les gènes de même variance (Delmar et al., 2005).

Conseil : j'ai tendance à utiliser le fichier VM-CompleteListOfGenes.txt.
Je l'ouvre sous excells en transformant les points en virgules puis je l'enregistre en « .xls ».
Ensuite je classe les gènes selon la colonne BH, les premiers gènes sont donc ceux qui sont déclarés différemment exprimés, le nombre dépendra du seuil que vous avez fixé.
Signification des colonnes de ces fichiers :

- les n premières colonnes sont le moyen d'identifier chaque gènes
- A = intensités moyennes de chaque lame
- Delta i : log2(Cond1/Cond2) sur la lame i
- Delta : log2(Cond1/Cond2) moyen sur l'ensemble des lames
- VarianceByGene : variance par gène
- NbObs : nombre d'observations non manquantes
- tau i : probabilité d'appartenir à chacun des groupes déterminés par le modèle de mélange sur les variances
- VarianceVM : variance issue de la méthode varmixt (groupe de gènes de même variance, Delmar et al., 2005)
- StatOfTestVM : statistique de test de la méthode varmixt
- PValueVM : p-value brute VM
- bonferroni : p-value corrigée selon la méthode de bonferroni
- BH : p-value corrigée selon la méthode de Benjamini-Hochberg

Abstract

Bovine mammary gland produces daily high amount of milk, equivalent to 10% of the animal weight. Animal feeding, genetics and terminal differentiation process of the mammary tissue can impact significantly animal productivity and milk composition. The intense biosynthetic and secretion activities of mammary gland need a large number of genes which expression can be analysed simultaneously with microarrays. In this work, we used gene expression profiling to establish gene networks that participate to the terminal differentiation process of the mammary tissue and that explain milk composition changes due to K232A polymorphism in the bovine *DGAT1* gene.

At first, we followed, using 22K bovine oligoarrays, gene expression profiles of caprine mammary tissue during pregnancy. Our results determine 3 major stages in the terminal differentiation process: (i) activation of functional differentiation before the first third of the pregnancy time ; (ii) tissue remodelling in which the secretory tissue is differentiating and proliferating in the mammary space – this phase is characterized by the expression of immune response genes ; (iii) acquisition of the secretory phenotype delimited by lipid secretion activity at the parturition.

Then, to determine cellular mechanisms underlying changes of milk composition with the non conservative K232A mutation occurring in the bovine *DGAT1* gene, we have compared, using microarray, gene expression profiles from mammary tissue biopsies of cows reared in well defined and controlled experimental conditions. To each homozygous cow AA/AA (K232) corresponded its full-sib (n=4) or half-sib (n=2) homozygous GC/GC (A232), at the same lactating stage and day in lactation. Our results confirmed a significant reduction of milk fat content (41.6 vs. 51.6 g/kg) and of unsaturated-middle chain fatty acid content (C14:1, C16:1), and a significant increase in saturated middle chain (C14:0) and in unsaturated-long chain (C18:1 *cis11*, C18:1 *cis12, C18:2 n-6*, C18:3 *n-3*) fatty acid content, for the GC/GC (A232) compared to AA/AA (K232) genotype. In addition, milk fat globules were significantly smaller. Gene expression profiling performed on this new data set revealed up-regulation of genes involved in lactose, lipid and protein biosynthesis for GC/GC (A232) individuals. Genes of the lipid biosynthesis up-regulated with this genotype partially explain observed differences in fine milk fat composition by giving alternative pathways of diacylglycerol metabolism.

Résumé

La glande mammaire d'une vache laitière haute productrice, produit quotidiennement une quantité de lait équivalente à 10% du poids de l'animal. L'alimentation, la génétique et le processus de différenciation terminale du tissu mammaire peuvent impacter significativement la productivité de l'animal et la composition du lait. L'activité intense de biosynthèse et de sécrétion de la glande mammaire fait intervenir un large répertoire de gènes dont l'expression peut aujourd'hui être analysée de façon globale et simultanée grâce à des outils tels que les puces à ADN. C'est ce type d'approche que nous avons mis en œuvre, dans ce travail de thèse, afin d'établir les réseaux de gènes qui participent au processus de différenciation terminale du tissu mammaire et qui permettent d'expliquer les variations de composition du lait observées avec la mutation K232A au locus *DGAT1* bovin.

Pour ce faire, nous avons utilisé chez la chèvre un dispositif en boucle qui a permis d'analyser, au moyen d'un répertoire de 22 000 sondes oligonucléotidiques bovines (22K), l'évolution du profil d'expression génique du tissu mammaire au cours de la gestation. Nous avons ainsi pu montrer que ce profil subit au cours du développement de l'organe et de sa différenciation terminale des changements profonds qui définissent 3 étapes majeures : i) le déclenchement de la différenciation fonctionnelle qui intervient avant le premier tiers de la gestation ; ii) un remodelage du tissu au cours duquel les structures qui produiront le lait (acini) se mettent en place et prolifèrent pour envahir progressivement l'organe - cette phase est marquée par l'expression de gènes de la réponse immunitaire ; iii) l'acquisition du phénotype sécrétoire, matérialisée par le déclenchement de la synthèse lipidique à la parturition.

Afin de déterminer les mécanismes cellulaire sous-jacents aux variations de composition du lait observées avec la mutation K232A au locus *DGAT1*, nous avons entrepris la comparaison du transcriptome de tissu mammaire bovin à partir d'un dispositif expérimental génétiquement bien défini et contrôlé. A chaque animal de génotype GC/GC (A232), correspondait soit sa pleine sœur (n=4), soit sa demi-sœur (n=2) de génotype AA/AA (K232), au même numéro de lactation et au même nombre de jours de lactation. Nos résultats ont confirmé une baisse significative du TB (41.6 vs. 51.6 g/kg) et des AG à chaîne moyenne et insaturés (C14:1 et C16:1) et une augmentation des AG à chaîne moyenne et saturée (C14:0) et longue et insaturée (C18:1 *cis11*, C18:1 *cis12, C18:2 n-6*, C18:3 *n-3*) avec le variant GC/GC (A232) en comparaison au variant AA/AA (K232). Les globules gras étaient également plus petits avec ce variant. L'analyse du transcriptome à partir d'ARN extraits des biopsies de tissu mammaire a montré que les gènes spécifiant les acteurs des voies de synthèse du lactose, des lipides et des protéines étaient sur-exprimés chez les individus de génotype GC/GC (A232). La liste des gènes de la biosynthèse des lipides sur-exprimés avec ce génotype expliquent, pour partie, les différences de composition fine en acides gras observées en suggérant des voies alternatives du métabolisme du diacylglycérol.