

Université
de Toulouse

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :
PATOGEN

Présentée et soutenue par :
Cécile BONNEFONT

le : lundi 14 novembre 2011

Titre :
Analyses génomiques fonctionnelles de la résistance aux mammites :
études de deux lignées divergentes de brebis
sélectionnées sur la concentration cellulaire du lait

Ecole doctorale :
Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :
UR631 INRA SAGA / UMR1225 INRA-INPT-ENVT IHAP

Directeur(s) de Thèse :
Christèle ROBERT-GRANIE
Gilles FOUCRAS (IHAP)

Rapporteurs :
Johann DETILLEUX
Eve DEVINOY

Membre(s) du jury :
Johann DETILLEUX
Eve DEVINOY
Gilles FOUCRAS
Pascal MARTIN
Pascal RAINARD
Rachel RUPP

REMERCIEMENTS

Une thèse, c'est du travail, beaucoup de travail, mais c'est aussi des rencontres. Je voudrais remercier toutes les personnes qui ont croisé mon chemin à un moment donné et qui m'ont aidé, par leurs conseils avisés, par leurs encouragements ou par leur bonne humeur.

Mon travail de doctorat a été réalisé entre la Station d'Amélioration Génétique des Animaux (SAGA) et l'UMR Interactions Hôtes-Agents Pathogènes (IHAP). Il a été financé par les départements de génétique animale (GA) et de santé animale (SA) de l'INRA. A ce titre, je tiens à remercier Didier Boichard et Denis Milan, ancien et actuel chefs du département GA, Gilles Aumont et Thierry Pineau, ancien et actuel chefs du département SA pour avoir accepté le co-financement de ma thèse. Je remercie aussi Christèle Robert-Granié et François Schelcher, directeurs respectifs de la SAGA et de l'IHAP pour m'avoir accueillie dans leur laboratoire, ainsi que toutes les personnes des deux unités pour leur convivialité.

Je remercie tous les membres du jury et particulièrement Johann Detilleux, professeur dans le groupe de Génétique Quantitative, de la faculté de médecine vétérinaire de l'université de Liège, pour avoir accepté d'être rapporteur de ma thèse et Eve Devinoy, directrice de recherche 1 de l'unité INRA Génomique et Physiologie de la Lactation (UR1196 GPL) qui m'a fait l'honneur d'être à la fois présidente du jury et rapporteur de ma thèse. Je remercie également Pascal Martin, ingénieur de recherche 2 dans l'unité ToxAlim (UMR1331 INRA-INPT-ENVT) pour ses conseils avisés sur les analyses du transcriptome et Pascal Rainard, directeur de recherche 2 dans l'unité INRA Infectiologie Animale et Santé Publique (UR1282 IASP) pour ses réflexions sur le fonctionnement intégré des cellules de la glande mammaire lors d'infections.

Je remercie également les membres du comité de pilotage de ma thèse Pascal Martin, Philippe Mulsant, Claire Rogel-Gaillard et Magali San Cristobal pour leurs précieux conseils sur les analyses de puce à ADN et pour leurs réflexions sur mes travaux de thèse. Je remercie chaleureusement Mehdi Toufeer, ancien doctorant de l'IHAP, qui m'a bien aidé dans la réalisation des analyses de laboratoire et qui m'a permis de mener à bien l'ensemble des projets. Je remercie aussi Pascal Rainard et son équipe avec qui j'ai collaboré pour les études des cellules épithéliales mammaires.

Je suis très reconnaissante à Rachel Rupp, Gilles Foucras et Christèle Robert-Granié, qui m'ont guidé tout au long de ma thèse. Et même si leurs points de vue et leurs objectifs n'étaient pas toujours identiques, et même si j'ai parfois rencontré quelques difficultés pour trouver ma place, c'était une richesse de travailler avec eux, et petit à petit, j'ai pu tracer mon petit bout de chemin entre la génétique quantitative et l'immunologie. Rachel, Gilles et Christèle je vous remercie chaleureusement pour avoir partagé avec moi vos connaissances (scientifiques, méthodologiques, humaines), pour m'avoir orientée pendant ces trois années, et m'avoir soutenue lorsque j'en avais besoin.

Je remercie aussi le personnel des plateformes GenoToul qui m'ont aidé lors des différentes étapes, Lidwine Trouilh pour les hybridations des puces à ADN, Olivier Bouchez pour les analyses de PCR à grande échelle et Philippe Bardou de Siganae pour le dépôt des données transcriptomiques dans GEO. Je remercie Magali San Cristobal, Laurence Liaubet,

Gwenola Tosser-Klopp, Christèle Robert-Granié et les autres membres du groupe de travail « Biopuce », ainsi que Sébastien Déjean, Pascal Martin et les autres membres du groupe de travail « Biostat » pour m'avoir formée, aidée et conseillée dans l'analyse des données transcriptomiques et toujours dans la bonne humeur.

Je remercie Guillaume Sallé, Guillaume Balloche, Simon Teyssède, Carole Moreno, Gwenola Tosser-Klopp et Julie Demars pour avoir partagé avec moi leurs connaissances sur les analyses de polymorphisme d'ADN et m'avoir fait bénéficier de certains de leurs outils. Je remercie aussi l'équipe d'Alessandra Stella du Parco Technolgico Padano avec qui j'ai eu l'occasion de collaborer, et principalement Bouabid Badaoui pour la méta-analyse des données transcriptomiques et Francesco Strozzi pour l'annotation de la puce à ADN ovine.

Et comme sans les brebis, je n'aurai pas eu de sujet d'étude, je voudrais remercier tout le personnel de l'unité expérimentale de La Fage qui participe à l'élevage du troupeau, et spécialement à Marie-Rose Aurel et Ophélie Duvallon ainsi que Charlotte Allain qui fait le lien entre La Fage et la SAGA. Je remercie également le personnel de l'Ecole Nationale Vétérinaire de Toulouse qui s'occupe des brebis présentes sur le campus.

Je remercie aussi Cécile Caubet et Christian Tasca pour leur aide en laboratoire, Guillaume Tabouret et Séverine Boullier pour leurs discussions autour des mammites ainsi que Emmanuelle Totain et Sokunthea Top pour leur bonne humeur.

Merci au petit groupe de jeunes sagassiens que j'ai appris à connaître et avec qui j'ai partagé de très bons moments. Merci à Carine, Simon, Guigui, Julien, Hugues, Mila, Chloé, Charlotte, Momo, Cyrielle, Guillaume. Un merci spécial à Libertat.

Et puis, il y a la vie en dehors du labo et je voudrais sincèrement remercier toutes les personnes qui gravitent dans ma petite sphère et qui me permettent d'oublier la fatigue, d'évacuer le stress et d'être heureuse. D'abord un grand merci à ma famille, à mes parents qui m'ont montré un bon modèle et qui me racontent de jolies histoires sur les petits animaux de la forêt ; à mes frérots Stéphane et Michel qui m'emmènent découvrir de super paysages en montagne en hiver comme en été ; à ma cantinière préférée du mercredi midi, j'ai nommée Domi qui, avec ma mamie et les deux petites demoiselles Noä et Ellie, me rend toujours le sourire.

Et puis un grand merci à ma famille « ultimate frisbee » à Lisou, Kaouet, Céd, JG, Vévert, je ne peux pas tous les citer, mais un merci particulier au petit groupe des motivés du lundi soir avec le fameux « Physique – Tartiflette ». Vive les BTR !!

Merci à la joyeuse bande auzevilloise que je suis toujours ravie de retrouver quelque soit le lieu. Merci à Céline, Alice, Sarah, Pascalou, Josépha et les autres ensatiens pour tous ces bons moments.

Et puis, je finirai par un merci spécial à Ludo, qui m'a bien aidé à gérer mes coups de stress et que je remercie également pour partager ma petite vie au quotidien.

J'ai aussi une petite pensée pour ceux qui m'ont quitté pendant ces trois dernières années : Arlette, Roger et Nelly.

RESUME

Les mammites sont des inflammations de la mamelle provoquées principalement par des bactéries. Elles représentent un problème majeur en élevage car elles entraînent d'importantes pertes économiques pour les filières de ruminants laitiers. Elles sont caractérisées par de fortes augmentations de la concentration des cellules somatiques dans le lait. Le score de cellules somatiques (SCS) est très bien corrélé à la présence d'une infection mammaire. Il peut donc être utilisé en sélection pour améliorer la résistance aux mammites. Pour comprendre les mécanismes mis en place lors d'une sélection sur les SCS, deux lignées divergentes de brebis ont été produites à partir de géniteurs ayant des valeurs extrêmes de l'index SCS.

Le principal objectif de ce travail est d'identifier et de comprendre les mécanismes qui confèrent une plus grande résistance ou sensibilité aux mammites provoquées par des staphylocoques. Nos travaux ont porté principalement sur des analyses transcriptomiques de trois types cellulaires majeurs, après contact avec des staphylocoques : les cellules inflammatoires du lait recueillies après infection, qui sont principalement composées de neutrophiles, les cellules dendritiques, comme cellules présentatrices d'antigène et les cellules épithéliales mammaires, première barrière de défense et qui sont les plus nombreuses sur le site de l'infection. Nous avons montré que la migration des cellules immunitaires et les processus inflammatoires se traduisent par des activations de voies différentes chez les brebis des deux lignées divergentes. Nous avons aussi identifié des gènes fonctionnels candidats (*ahr*, *tp53*, *sc5* et *tmem87b*) pour expliquer la différence de sensibilité aux mammites. En parallèle, nous avons étudié l'association entre le polymorphisme de l'ADN et la présence d'abcès mammaires liés aux mammites dans un dispositif cas-contrôle visant à détecter des gènes à effet majeur. Les études ont permis de mettre en évidence une zone chromosomique d'intérêt située sur le chromosome OAR5.

Les deux approches de génétique moléculaire menées dans ces études, à la fois l'analyse du transcriptome et celle du polymorphisme du génome, ont permis d'identifier des gènes fonctionnels et positionnels candidats pour comprendre les mécanismes associés à la résistance aux mammites.

ABREVIATIONS

ADN : Acide Desoxyribonucléique
ADNc : ADN complémentaire
ARNm : Acide Ribo Nucléique messenger
BLUP : Best Linear Unbiased Prediction
BMDC : Bone Marrow derived Dendritic Cells
CAEV : Arthrite Encéphalite Caprine à Virus
CCS : Concentration de cellules somatiques
CD : Cellules dendritiques
CEM : Cellules épithéliales mammaires
CMH : Complexe majeur d'histocompatibilité
CNV : Copy Number Variation, variation du nombre de copie
CR4 : Complement Receptor 4
CSF : Colony Stimulating Factor
CSL : Cellules somatiques du lait
CTIG : Centre de Traitement de l'Informatique Génétique
ENVT : Ecole Nationale Vétérinaire de Toulouse
FC : Fold-Change
FCEL : France Conseil Elevage
GMCSF : Granulocyte Macrophage Colony Stimulating Factor
PNN : Poly-Nucléaire Neutrophile
GSEA : *Get Set Enrichment Analysis*
HKG : House Keeping Genes : gènes de ménage
IA : Insémination Artificielle
Ig : Immunoglobulines
IL : Interleukine
IMI : Infection Intra-Mammaire
INRA : Institut National de la Recherche Agronomique
LA : Linkage Analysis
LD : Linkage Disequilibrium
LDLA : Linkage Disequilibrium Linkage Analysis
LPS : Lipopolysaccharide
LTA : Acide Lipotéicoïque
MC : Mammite Clinique

MCSF : Macrophage Colony Stimulating Factor
MIAME : Minimum Information About a Microarray Experiment
MKK : MAPK kinase
MRSA : Methicillin-Resistant *Staphylococcus aureus*
MSC : Milk Somatic Cell, cellules somatiques du lait
NADPH : Nicotinamide Adenine Dinucleotide Phosphate-oxidase
NK : Natural Killer cell, cellules naturelles tueuse,
PAMP : Pattern Associated Molecular Pathogen
PBMC : Peripheral Blood Mononuclear Cell
QTL : Quantitative Trait Locus
RNAseq : séquençage d'ARN
ROS : Reactive Oxygen Specie
RT-qPCR : quantitative Reverse Transcription Polymerase Chain Reaction
Sa : S. aureus : *Staphylococcus aureus*
SAGA : Station d'Amélioration Génétique des Animaux
SAM : Sélection Assistée par Marqueurs
SCN : Staphylocoque à Coagulase Négative
SCP : Staphylocoque à Coagulase Positive
SCS : Score de cellules somatiques, Somatique Cell Score
Se : S. epidermidis : *Staphylococcus epidermidis*
SIG : Système d'Information Génétique
SNP : Single Nucleotide Polymorphism
Sp : supernatant : surnageant
TAQ : *Thermophilus aquaticus*
TFBS : *Transcription Factor Binding Site*
Th : T helper
TLR : Toll Like Receptor
TNFa : Tumor Necrosis Factor α
UE : Unité Expérimentale

SOMMAIRE

Remerciements	1
Résumé	3
Abréviations	4
Sommaire.....	6
Table des illustrations.....	10
Introduction Générale.....	13

Chapitre I : Les mammites des ruminants laitiers et les objectifs	17
-------------------------------------------------------------------------	----

1	Les infections intra-mammaires : définition, causes et évolution	19
1.1	Mammites cliniques et subcliniques	19
1.2	Etiologie des mammites : des causes multiples	20
1.3	Dynamique des infections : contamination et évolution clinique	22
1.4	Critères de détection des infections intra-mammaires	22
1.4.1	Le diagnostic bactériologique	22
1.4.2	La détection des cas cliniques	23
1.5	Critères indirects et dépistage des infections intra-mammaires	24
1.6	Conséquences biologiques	26
2	Les mécanismes généraux de protection contre les infections mammaires chez les ruminants laitiers	27
2.1	Les mécanismes de défense naturelle.....	27
2.2	Les mécanismes de défense immunitaire.....	28
2.2.1	Immunité innée de la glande mammaire	28
2.2.1.1	Immunité cellulaire.....	29
2.2.1.1.1	Récepteurs Toll-Like et défenses de la glande mammaire.....	29
2.2.1.1.2	Les neutrophiles.....	30
2.2.1.1.3	Les macrophages	33
2.2.1.1.4	Les cellules dendritiques	33
2.2.1.1.5	Les cellules naturelles tueuses.....	33
2.2.1.1.6	Les cellules épithéliales mammaires	34
2.2.1.2	La fraction soluble de l'immunité innée.....	34

2.2.1.2.1	Les peptides antimicrobiens	34
2.2.1.2.2	Le système du complément	34
2.2.2	La réponse immunitaire acquise de la glande mammaire	35
2.2.2.1	Les lymphocytes et les défenses mammaires	35
2.2.2.2	Les cellules B et les immunoglobulines	35
2.2.2.3	Le rôle des cytokines dans la glande mammaire	36
2.3	Variation du transcriptome lors d'infections mammaires	36
3	Lutte contre les mammites : des méthodes curatives et préventives	38
3.1	Méthodes de lutte classique.....	38
3.1.1	Des traitements curatifs : les antibiotiques.....	38
3.1.2	Des méthodes préventives	39
3.1.2.1	Hygiène et gestion du troupeau	39
3.1.2.2	Vaccination.....	39
3.2	L'amélioration génétique	40
3.2.1	Qu'est-ce que la résistance et comment la mesurer ?.....	40
3.2.2	Les paramètres génétiques	41
3.2.3	Génétique quantitative et sélection.....	44
3.2.3.1	Evaluation génétique : calcul de l'index SCS	45
3.2.3.2	Utilisation de l'index SCS dans les programmes d'amélioration génétique	45
3.2.4	Les apports de la génétique moléculaire	46
3.2.4.1	Le polymorphisme de l'ADN.....	46
3.2.4.2	Les QTL associés à la résistance aux mammites.....	47
3.2.4.3	Utilisation de l'information moléculaire	48
4	Objectifs de la thèse	50

Chapitre II : Matériels et méthodes	53
-------------------------------------------	----

1	Un modèle d'étude original : les lignées divergentes.....	55
1.1	Protocole de sélection divergente.....	55
1.1.1	Production des brebis des lignées divergentes	55
1.1.2	Des différences de concentrations de cellules somatiques.....	56
1.2	Résistance aux infections intra-mammaires dans les lignées divergentes ...	56

1.2.1	Suivi d'infections naturelles.....	56
1.2.2	Suivi d'infections expérimentales	57
2	Analyses des différences d'origine génétique	59
2.1	L'analyse des gènes exprimés	59
2.1.1	Quelques méthodologies d'analyse	59
2.1.1.1	Quantification des ARN par réaction de polymérisation en chaîne ..	59
2.1.1.2	Quantification des ARN par puce à ADN	61
2.1.1.2.1	Nature et définitions d'une puce à ADN	61
2.1.1.2.2	Hybridation des échantillons d'ARNm	63
2.1.1.2.3	Nettoyage et normalisation des données	64
2.1.2	Etude de la variabilité biologique dans l'expression des gènes	66
2.1.2.1	Identification de gènes différemment exprimés	67
2.1.2.1.1	Analyse de variance avec des modèles linéaires	67
2.1.2.1.2	Correction sur les tests multiples.....	67
2.1.2.1.3	Mesure des différences d'expression entre les conditions biologiques	68
2.1.2.1.4	Cas particulier des données de RT-qPCR	68
2.1.2.2	Interprétation des listes de gènes	69
2.1.2.2.1	Les annotations de la puce ovine Agilent.....	69
2.1.2.2.2	Analyses de groupes de gènes	69
2.1.2.3	Analyse d'enrichissement.....	73
2.1.3	Comment confirmer les résultats ?.....	73
2.2	Etude du polymorphisme de l'ADN.....	74
2.2.1	Un outil de choix pour l'étude du polymorphisme chez l'ovin : la puce SNP 54K d'Illumina.....	74
2.2.2	Les méthodes d'étude du polymorphisme de l'ADN.....	74
2.2.3	Vers l'identification des gènes et mutations causales	76

Chapitre III : Résultats et discussions partielles.....	79
---------------------------------------------------------	----

Partie 1 : Analyse transcriptomique des cellules somatiques du lait chez des brebis de la lignée résistante et de la lignée sensible aux mammites après infection expérimentale par *Staphylococcus epidermidis* et *Staphylococcus aureus*

1	Contexte et objectifs.....	81
2	Résumé	82

3	Article 1.....	83
4	Discussion.....	101
Partie 2 : Analyse transcriptomique des cellules dendritiques dérivées de la moelle osseuse chez des brebis de la lignée résistante et de la lignée sensible aux mammites après stimulation <i>in vitro</i> par <i>Staphylococcus aureus</i>		
1	Contexte et objectifs.....	105
2	Résumé	106
3	Article 2.....	107
4	Discussion	123
Partie 3 : Cellules épithéliales mammaires et résistance aux mammites : analyse transcriptomique après infection expérimentale <i>in vitro</i> par <i>Staphylococcus aureus</i> ou stimulation par son surnageant de culture		
1	Contexte et objectifs.....	125
2	Résumé	125
3	Article 3.....	127
4	Discussion	163
Partie 4 : Méta-analyse des données transcriptomiques réalisées à partir des cellules de brebis sélectionnées sur la résistance aux mammites.....		
1	Contexte et objectifs.....	165
2	Résumé	168
3	Article 4.....	171
4	Discussion	191
Partie 5 : Etude du polymorphisme de l'ADN dans les populations de brebis CCS+ et CCS-.....		
1	Contexte et objectifs.....	195
2	Matériel et méthode.....	195
3	Résultats	199
4	Discussion	205
Chapitre IV : Discussion et perspectives.....		211
Annexes		225
Bibliographie		233

TABLE DES ILLUSTRATIONS

Légende des figures

Figure 1 : Maturation des granulocytes	31
Figure 2 : Diapédèse des neutrophiles à travers la couche de cellules endothéliales... 31	
Figure 3 : Score moyen des examens cliniques des mamelles et excréments bactériennes après inoculation de 10^3 ufc <i>S. epidermis</i> et de <i>S. aureus</i>	58
Figure 4 : Courbes de suivi de la quantité d'ADN au cours de la RT-qPCR	60
Figure 5 : Courbes de dissociation des ADN double brin en ADN mono brin	60
Figure 6 : Histogramme du nombre de publications utilisant la technologie microarray par an depuis 1990.....	62
Figure 7 : Distribution des intensités de fluorescence de tous les spots d'un microarray selon la normalisation des données	65
Figure 8 : MAplot des intensités de fluorescence selon la normalisation des données de microarray	65
Figure 9 : Classification hiérarchique des gènes avec un heatmap	70
Figure 10 : Exemples de deux réseaux de gènes obtenus avec IPA.....	72
Figure 11 : Dénombrement des neutrophiles sanguins dans les six jours après l'inoculation intra-mammaire de <i>S. aureus</i>	102
Figure 12 : Diagramme de Venn entre les trois listes de gènes différentiels entre lignées obtenus avec les CSL, CD et CEM.....	167
Figure 13 : Réseau de gènes obtenu avec Ingenuity (IPA) qui regroupe les gènes différentiels dans chacune des analyses (CSL, CD et CEM).....	168
Figure 14 : Nombre de SNP et espacement moyen entre deux SNP par OAR	198
Figure 15 : Significativité de tous les SNP dans l'analyse des lignées divergentes de brebis.....	200
Figure 16 : Significativité des SNP dans l'analyse des lignées divergentes de brebis sur les chromosomes OAR7, OAR8 et OAR10 qui contiennent au moins un SNP significatif. .	201
Figure 17 : Significativité de tous les SNP dans l'analyse binaire et catégorielle des abcès	202
Figure 18 : Significativité des SNP dans l'analyse binaire des abcès sur les chromosomes OAR1, OAR2, OAR5, OAR7, OAR8, OAR21 et OAR24 qui contiennent au moins un SNP significatif	203
Figure 19 : Significativité des SNP dans l'analyse binaire des abcès sur les chromosomes OAR5, OAR8, OAR16, OAR21 et OAR24 qui contiennent au moins un SNP significatif.....	204
Figure 20 : Voies de signalisation à partir des TLR et d'autres récepteurs.....	215

Légende des tableaux

Tableau 1 : Réservoirs et fréquence relative des microorganismes responsables de mammites cliniques et de mammites subcliniques	21
Tableau 2 : Règles de prédiction des IMI basées sur les CCS mensuels	25
Tableau 3 : Modifications physico-chimiques du lait causées par les mammites	27
Tableau 4 : Récapitulatif des PRR et de leurs principaux ligands	29
Tableau 5 : QTL pour la résistance aux mammites chez les brebis laitières.....	50
Tableau 6 : Caractéristiques des brebis CCS+ et CCS-.....	56
Tableau 7 : Concentrations de cellules somatiques et fréquences des mammites chez les brebis CCS+ et CCS-	57
Tableau 8 : Effectifs des brebis dans chaque analyse de polymorphisme d'ADN.....	196
Tableau 9 : Description des SNP significatifs dans l'analyse des lignées divergentes de brebis	200
Tableau 10 : Description des animaux présents dans l'analyse binaire des abcès	201
Tableau 11 : Description des SNP significatifs dans l'analyse binaire des abcès	202
Tableau 12 : Description des animaux présents dans l'analyse catégorielle des abcès	204
Tableau 13 : Description des SNP significatifs dans l'analyse catégorielle des abcès	205
Tableau 14 : Description des SNP situés à proximité du SNP significatif sur OAR5.. ..	205
Tableau 15 : Fréquence génotypiques des 4 SNP situés à proximité du SNP significatif sur OAR5 dans l'analyse binaire des abcès.. ..	206
Tableau 16 : Fréquence génotypiques des 4 SNP situés à proximité du SNP significatif sur OAR5 dans l'analyse catégorielle des abcès.....	206
Tableau 17 : Description des clusters de gènes différemment exprimés.....	208

TABLE DES ANNEXES

Annexe 1 : Poster WCGALP, 2010.....	227
Annexe 2 : poster EEAP 2011.....	228
Annexe 3 : Répartition des SNP sur chaque chromosome et espacement moyen.....	229
Annexe 4 : Liste des communications.....	230

INTRODUCTION GENERALE

Les mammites représentent un problème sanitaire et économique majeur en élevage de ruminants laitiers. Pour pallier leurs effets, la plupart des programmes de sélection de bovins laitiers en vigueur à travers le monde prend en compte les concentrations de cellules somatiques (CCS) et éventuellement les cas de mammites cliniques. Fin 2008, peu de données sur les effets concrets *a posteriori* d'une sélection sur les CCS étaient disponibles.

Une expérience de sélection divergente de brebis à partir des index SCS extrêmes a permis d'obtenir deux lignées avec de grandes différences phénotypiques. Les brebis avec des concentrations cellulaires du lait régulièrement plus élevées sont plus sensibles aux infections intra-mammaires (IMI) que les brebis avec des concentrations cellulaires plus faibles. Ces différences de résistance se caractérisent par des fréquences de mammites cliniques, subcliniques et d'abcès plus élevées chez les brebis de la lignée sensible que chez les brebis de la lignée résistante. Pour autant, les mécanismes mis en place lors de la sélection sur le critère CCS ne sont pas connus et il est probable que les différences génétiques entre les deux lignées soient polygéniques.

Ce travail de thèse a pour vocation d'apporter des connaissances sur le déterminisme génétique de la sélection sur les CCS et sur les mécanismes de réponse de l'hôte à une IMI, en valorisant le modèle des lignées divergentes. Les analyses se basent sur l'étude du transcriptome et du polymorphisme génétique avec des puces à ADN (appelées microarray dans le manuscrit) et des puces permettant d'identifier des polymorphismes d'un seul nucléotide (appelées puces à SNP – *Single Nucleotide Polymorphism*), respectivement. Ces deux supports analytiques spécifiques à l'espèce ovine ont été développés récemment et nous ont permis d'investiguer les mécanismes ou les gènes qui expliquent les différences de résistance aux IMI entre les deux lignées de brebis.

Le Chapitre 1 établit le contexte de mes études à partir de données bibliographiques sur les mammites, en montrant pourquoi les infections intra-mammaires (IMI) sont considérées comme la première cause de maladie en élevage laitier. Ce chapitre est traité sous trois aspects : quelles sont les causes des mammites ?, quelles sont leurs conséquences ? et comment diminuer leur impact ? De nombreuses informations proviennent d'études réalisées sur la vache laitière, mais certaines spécificités de la brebis et de la chèvre laitières sont abordées. Afin de comprendre l'immunité mammaire, les principaux mécanismes immunitaires de réponse aux infections à staphylocoque chez les ruminants laitiers sont décrits brièvement. Puis les différents moyens de lutte contre les mammites sont abordés d'un point de vue curatif mais aussi préventif, en développant particulièrement l'amélioration génétique. A la fin de ce chapitre, les objectifs de mon travail de thèse sont présentés.

Le Chapitre 2 récapitule les principales informations méthodologiques nécessaires à la réalisation des expériences décrites dans ce manuscrit. Dans un premier temps, les ressources animales sont présentées afin d'établir leurs caractéristiques et de positionner les objectifs de mon travail. Puis les principaux outils et méthodes utilisés pour les analyses du polymorphisme génétique et du transcriptome sont présentés synthétiquement.

L'ensemble des résultats obtenus est présenté dans le Chapitre 3 sous forme de partie. Chaque partie décrit une analyse différente et se découpe selon un modèle commun qui s'appuie sur un article en anglais (le statut de l'article, publié, soumis ou en préparation est précisé), à l'exception de la partie 5, qui est rédigée entièrement en français. L'article est précédé du contexte et des objectifs de l'expérience présentée, ainsi que d'un résumé en français des principaux résultats de chaque analyse. Il est suivi d'une discussion en français qui permet de reprendre les principales informations de l'article et de proposer de nouvelles hypothèses.

La Partie 1 traite des réponses des cellules somatiques inflammatoires du lait après inoculation de *Staphylococcus epidermidis* et de *Staphylococcus aureus* dans les deux lignées divergentes de brebis. Les Parties 2 et 3 présentent les réponses *in vitro* des cellules dendritiques et des cellules épithéliales mammaires, respectivement, après stimulation par *S. aureus*. Ces trois parties permettent d'établir des hypothèses sur les mécanismes cellulaires de réponse aux infections à *Staphylococcus* qui différencient les brebis de la lignée résistante de celles de la lignée sensible.

Ensuite, une méta-analyse de ces trois jeux de données permet de confirmer certaines hypothèses qui ont été formulées lors des analyses monocellulaires, et d'en générer de nouvelles. Elle permet aussi d'établir une liste de mécanismes et de gènes candidats fonctionnels pour déterminer l'origine génétique de la différence entre lignées. Cette synthèse est présentée dans la Partie 4.

Pour compléter l'étude génétique des lignées, des études du polymorphisme de l'ADN sur ces animaux ont été réalisées (Partie 5). Dans un premier temps, ces analyses portent sur les lignées divergentes de brebis, puis sur la formation des abcès mammaires qui symbolise des mammites chroniques. Ensuite, les résultats des analyses fonctionnelles et positionnelles sont confrontés pour mettre en évidence les principales caractéristiques des brebis de la lignée résistante et de celles de la lignée sensible aux IMI.

Le chapitre IV récapitule brièvement l'ensemble des résultats obtenus et les compare aux données publiées sur la résistance aux mammites d'origine génétique. Puis, quelques réflexions sur le travail effectué sont apportées ainsi que des perspectives pour la poursuite de ce projet.

CHAPITRE I : LES MAMMITES DES RUMINANTS LAITIERS ET LES OBJECTIFS

Les mammites sont définies comme des conditions inflammatoires de la glande mammaire en réponse à une blessure, qui servent à détruire et neutraliser les agents infectieux et à promouvoir la guérison et le retour à un fonctionnement normal de la glande mammaire (National Mastitis Council 1996).

La fréquence des mammites cliniques est généralement comprise entre 20 et 45%, chez les bovins. Il a été montré en Norvège qu'elle avait fortement augmenté entre 1978 et 1994 (de 13% vaches infectées en première lactation vs 28%) (Heringstad et al. 1999). Chez les petits ruminants, elle est bien plus faible (de l'ordre de 5%) (Bergonier *et al.* 2003). La fréquence des mammites subcliniques est bien plus élevée : entre 47 et 53% chez les bovins en France (Fabre et al. 1997; Faye et al. 1994) et entre 15 et 55% chez les petits ruminants (Paape et Capuco 1997), mais les données varient fortement d'une étude à l'autre.

Selon, la sévérité de l'infection et le nombre d'animaux malades au sein du troupeau, les IMI peuvent provoquer de fortes pertes économiques pour les éleveurs. Ces pertes reposent sur la diminution du volume de lait produit, la dévalorisation du prix de vente du litre de lait (pénalités dans le cadre du paiement du lait à la qualité), les cas cliniques (traitements, volume de lait non commercialisé, ...) et l'augmentation du taux de réforme des animaux. Au Royaume-Uni, l'ensemble des pertes liées aux mammites est estimé entre 106 et 373 millions de livres par an (Axford et al. 2000) et à 2 milliards de dollars aux Etats-Unis comme le rapporte Sordillo *et al.* (2002). Ainsi, les mammites sont classées au premier rang des maladies chez les bovins laitiers en regard de leurs conséquences économiques (Davies et al. 2009).

Ma thèse s'inscrit dans le cadre des mammites et elle a pour but de mieux comprendre les IMI. Dans ce premier chapitre, les mammites sont présentées à travers leurs causes, leurs conséquences et les méthodes qui permettent de les diagnostiquer. Ensuite, les mécanismes généraux de défense contre les IMI à des niveaux physiologiques, puis immunitaires sont présentés. Une synthèse bibliographique sur les mécanismes fins de réponse aux IMI est établie en se basant sur les analyses transcriptomiques qui ont été réalisées à la suite d'infection par des agents responsables de mammites. A la fin de ce chapitre, les méthodes de lutte contre les mammites sont présentées avec d'un côté les méthodes classiques (préventives et curatives) réalisées en élevage et de l'autre l'amélioration génétique en détaillant les paramètres génétiques des critères utilisés et la sélection sur ces critères.

1 Les infections intra-mammaires : définition, causes et évolution

1.1 Mammites cliniques et subcliniques

Les IMI sont provoquées par la pénétration d'un microorganisme dans la glande mammaire par le canal du trayon. Selon sa sévérité, on différencie la **mammite clinique** (entraînant une modification systématique de l'aspect du lait), de la **mammite subclinique** que l'on met en évidence *a posteriori*, grâce à une bactériologie du lait et à l'identification d'un recrutement cellulaire. Parfois les germes s'installent de façon durable dans la glande mammaire, on parle alors de **mammites chroniques**.

Différents symptômes peuvent avoir lieu lors d'une mammite clinique. Premièrement des symptômes locaux peuvent apparaître i) au niveau de la glande mammaire avec un gonflement de la mamelle, souvent accompagné de douleurs, une augmentation de la température et parfois une congestion (couleur rougeâtre) et ii) au niveau du lait avec la

formation de grumeaux, parfois accompagnés de liquide séro-hémorragique (mélange de sérum et de sang), voire de pus. Dans les cas les plus graves, des symptômes généraux apparaissent, ils sont dus à la libération d'IL1 β , IL6 et TNFa qui agissent sur le foie et le système nerveux central. Le foie peut alors sécréter des protéines de la phase aiguë et le système nerveux central peut provoquer de la fièvre. Cette élévation de la température corporelle est généralement accompagnée d'une accélération du rythme cardiaque et du rythme respiratoire.

1.2 Etiologie des mammites : des causes multiples

Les mammites sont provoquées par plus d'une centaine de microorganismes (Rinaldi et al. 2010; Watts 1988). Elles sont presque exclusivement d'origine infectieuse. Elles sont souvent liées à des traumatismes ou des agressions de la peau du trayon ou au niveau de la mamelle et parfois à des sténoses du canal du trayon (lésions liées à la machine à traire). Elles sont essentiellement provoquées par des **bactéries** et beaucoup plus rarement par des levures, champignons, parasites ou virus. Une mamelle saine est considérée comme stérile ; elle n'héberge donc pas de flore « normale ».

Les bactéries pathogènes peuvent être classées selon la **gravité des infections** qu'elles occasionnent (Tableau 1). Les bactéries responsables de mammites sévères sont qualifiées de **pathogènes majeurs**. Dans cette catégorie, on regroupe habituellement les staphylocoques à coagulase positive (SCP dont *Staphylococcus aureus*), les streptocoques, les entérobactéries (dont *Escherichia coli*) et les entérocoques. Parmi ces pathogènes, certains provoquent fréquemment des épisodes cliniques suraigus ou aigus (entérobactéries et *Streptococcus uberis*) alors que d'autres (comme *S. aureus*) sont surtout responsables de mammites chroniques. On qualifie de **pathogènes mineurs** les microorganismes qui produisent des infections modérées, principalement de type subclinique. Les staphylocoques à coagulase négative (SCN) et *Corynebacterium bovis* en sont les principaux représentants.

Les agents pathogènes peuvent également être distingués en fonction de leur **principal réservoir** (Tableau 1). Les staphylocoques, *Streptococcus agalactiae* et *Streptococcus dysgalactiae* sont présents préférentiellement dans la **mamelle** et sur la peau des trayons alors que les entérobactéries et les entérocoques se situent préférentiellement dans l'**environnement** (litière, fourrage, air). D'autres germes, comme *Streptococcus uberis*, peuvent occuper les deux niches.

Des virus peuvent être impliqués dans le déclenchement des mammites, soit en causant des lésions du trayon et ainsi en favorisant la contamination par d'autres pathogènes, soit en ayant une action immunosuppressive (Barkema et al. 2009; Wellenberg et al. 2002).

Tableau 1 : Réservoirs et fréquence relative des microorganismes responsables de mammites cliniques et de mammites subcliniques (d'après Poutrel 1985, Bergonier et al. 1997)

	<i>Staphylococcus aureus</i>	<i>Streptococcus agalactiae</i>	<i>Streptococcus dysgalactiae</i>	<i>Streptococcus uberis</i>	Entérobactéries Entérocoques (1)	<i>Staphylocoques à goagulase négative</i>	<i>Coryné-bactéries</i>
Réservoir principal	Mamelle/trayon	Mamelle/trayon	Mamelle/trayon	Mixte	Environnement	Mamelle/trayon	Environnement
Mammites cliniques							
Bovins	+	+++	+++	+++	++	--	--
Ovins	+++	-	-	+/-	--	++	--
Caprins	+++	-	-	-	--	++	--
Mammites subcliniques							
Bovins	+++	++	++	++	-	++	-
Ovins	++	-	-	-	--	+++	--
Caprins	++	-	-	-	--	+++	--

⁽¹⁾ Entérobactéries = *Escherichia coli*, *Klebsiella sp.*, *Proteus ...* ;
Entérocoques = *Enterococcus faecalis*, *Enterococcus faecium*

Néanmoins, l'étiologie des mammites présente des **particularités selon l'espèce** animale. Les SCN, considérés comme des agents pathogènes mineurs chez la vache, ont un rôle avéré et non négligeable chez les petits ruminants (de Clermont 1992; Seegers et al. 1997; Bergonier et al. 1997; Rupp et al. 2009) (Tableau 1). Lors de mammites cliniques, les streptocoques, entérobactéries et *S. aureus* sont prédominants chez la vache, alors que chez les petits ruminants les SCN et *S. aureus* sont plus fréquents (Tableau 1). Lors de mammites subcliniques, les staphylocoques (*S. aureus* et SCN) sont présents chez les trois espèces de ruminants, alors que les streptocoques sont essentiellement retrouvés chez la vache (de Clermont 1992; Seegers et al. 1997; Bergonier et al. 1997; Rupp et al. 2009). Chez les petits ruminants, les mycoplasmes occupent une place non négligeable et peuvent être responsables de mammites chroniques asymptomatiques ou de mammites sévères se traduisant par une agalactie totale (de Clermont 1992). Le virus de l'arthrite encéphalique caprine (CAEV) est impliqué dans les mammites subcliniques et parfois dans des formes cliniques avec une forte induration du tissu mammaire (de Clermont 1992; Lerondelle et al. 1992).

Comme *S. aureus* et les SCN sont les principaux agents pathogènes responsables de mammites chez les brebis, nous concentrerons nos études sur les *Staphylococci* et principalement à *S. aureus*. Pour mieux comprendre les mécanismes de réponse aux IMI, il est important de considérer la dynamique d'infection des hôtes.

1.3 Dynamique des infections : contamination et évolution clinique

Les agents pathogènes pénètrent dans la mamelle par le canal du trayon, les infections bactériennes par voie endogène (hématophore) étant exceptionnelles. Puis ils se déplacent du trayon vers les alvéoles de la citerne mammaire où le lait est synthétisé (Rinaldi et al. 2010).

L'évolution de l'infection dépend de l'espèce pathogène et de ses facteurs de virulence (Harmon 1994; Pearson et Mackie 1979), de la gestion du troupeau par l'éleveur, ainsi que de l'efficacité des défenses de l'animal (Burton et Erskine 2003). Ainsi, trois issues sont possibles : i) la guérison spontanée, mais elle est peu fréquente, ii) la persistance de l'infection traduisant un équilibre entre la multiplication du microorganisme et les défenses de l'animal et iii) l'évolution défavorable de l'infection pouvant aboutir à la perte complète ou partielle de fonctionnalité de la glande, et dans les cas extrêmes, à la mort de l'animal.

Par ailleurs, les mammites présentent un caractère évolutif. Une mammite subclinique peut évoluer en forme clinique, chronique ou aiguë, et inversement (Poutrel 1985; de Clermont 1992).

Ainsi les IMI débutent dans le trayon puis se propagent dans la glande mammaire où elles peuvent prendre différentes formes. Pour lutter contre les IMI, la première étape est de pouvoir les diagnostiquer.

1.4 Critères de détection des infections intra-mammaires

Les IMI peuvent être détectées soit par un diagnostic bactériologique du lait qui permet l'identification d'un pathogène, soit par l'observation des signes cliniques.

1.4.1 Le diagnostic bactériologique

Le diagnostic bactériologique consiste à mettre en culture des échantillons de lait dans un milieu contrôlé pour faire des clones de l'agent pathogène. Ces clones sont soumis à

différents tests afin de les caractériser. Actuellement, des méthodes de réaction de polymérisation en chaîne (PCR) ou de pyroséquençage permettent d'identifier l'espèce du pathogène. Les analyses bactériologiques peuvent être complétées avec d'autres méthodes qui permettent de dénombrer les agents pathogènes et peuvent apporter des éléments d'information sur le niveau de contamination d'un animal.

Un suivi régulier des quatre quartiers ou des deux héli-mamelles (au moins une fois par mois (Miller 1984)) permet un bon suivi de l'évolution de l'infection. La détection répétée d'un même pathogène chez une vache marque la présence d'une infection chronique. L'alternance de bactériologie négative, positive et négative, révèle la présence d'un épisode infectieux qui a été résolu. Des bactériologies répétées négatives, alors que l'animal est exposé à des agents pathogènes, signent la présence d'un animal résistant.

Le diagnostic bactériologique répété de toutes les mamelles est considéré comme la meilleure mesure de détection des IMI (Miller 1984), car il permet de détecter les mammites cliniques, les mammites subcliniques et d'identifier l'agent pathogène responsable. Si le diagnostic bactériologique est réalisé peu de temps après le prélèvement, le vétérinaire peut adapter le traitement des animaux malades spécifiquement contre l'agent pathogène identifié.

Toutefois, si les intervalles de temps sont trop longs entre deux prélèvements, des résultats négatifs peuvent cacher des épisodes infectieux de courte durée. Il peut exister quelques biais à cette méthode liés à des erreurs d'identification de germe. Le diagnostic bactériologique est habituellement effectué dans des laboratoires agréés, ce qui requiert une logistique de prélèvement et de transport des échantillons et entraîne un coût non négligeable (environ 12-15 € / échantillon). Il est donc difficile d'envisager un diagnostic bactériologique des ruminants en routine (Pyorala 2003).

1.4.2 La détection des cas cliniques

Les mammites cliniques peuvent aussi être diagnostiquées directement par l'éleveur soit par l'observation et/ou la palpation de la mamelle (gonflement, rougeur, durcissement,...), soit par l'observation des premiers jets de lait avant la traite (grumeaux). Mais les symptômes varient d'un animal à l'autre et selon l'intensité de l'infection.

Elles ont l'avantage d'être détectées rapidement, ainsi les animaux malades peuvent être isolés du troupeau au moment de la traite et ils peuvent recevoir un traitement aussitôt. Mais leur diagnostic ne permet pas d'identifier l'agent pathogène en cause (même si des hypothèses peuvent être formulées). Cependant les examens cliniques du tissu mammaire et du lait ne permettent pas de détecter les mammites subcliniques qui, par définition, ne se traduisent pas par des signes cliniques, mais uniquement par la présence de germes.

Les mammites cliniques des vaches sont enregistrées dans les pays scandinaves à grande échelle depuis les années 1980. Dans ces pays, tous les cas traités sont enregistrés de façon systématique par les vétérinaires fonctionnaires et l'information est centralisée dans d'importants recueils de données, ce qui permet d'avoir une bonne qualité des données enregistrées. En France, les données sont enregistrées via les éleveurs qui notent sur leur cahier de suivi d'élevage les cas de mammites apparus, puis un technicien de France conseil élevage (FCEL) enregistre ces informations et les fait remonter vers le Système d'Information Génétique (SIG). Mais le recueil des informations directement par les éleveurs pose des problèmes de standardisation de l'identification d'un cas de mammite.

Chez les petits ruminants, les symptômes cliniques locaux peuvent être observés relativement facilement (Bergonier et al. 1997), compte tenu du moindre volume de la

mamelle et de la facilité de la palpation. Les mammites cliniques chroniques, caractérisées par des lésions et des abcès mammaires, peuvent donc également faire l'objet d'enregistrements (Bergonier et al. 1997). Depuis quelques années, les réformes causées par des mammites sont enregistrées dans la base de données ovine nationale laitière (SIEOL).

1.5 Critères indirects et dépistage des infections intra-mammaires

Contrairement aux méthodes directes, les méthodes indirectes ne permettent qu'une prédiction de l'infection, sans identification de l'agent pathogène, mais elles ont souvent l'avantage d'être mises en œuvre plus facilement.

Le principal critère indirect de détection des mammites repose sur les **concentrations de cellules somatiques** du lait (CCS).

Chez les ruminants laitiers, en absence d'IMI, les valeurs des CCS sont faibles. Même si les valeurs observées varient selon les études, elles sont généralement inférieures à 100 000 ou 300 000 cellules/ml chez la vache (Serieys 1985a; Harmon 1994; Laevens et al. 1997), inférieures à 100 000, 150 000 ou 300 000 cellules/ml chez la brebis (Morgante et al. 1996; Bergonier et al. 1994; Gonzalez-Rodriguez et al. 1995) et inférieures à 750 000 cellules/l chez la chèvre (de Crémoux et al. 1994; de Crémoux et al. 1996). Toutefois, la composition des cellules somatiques du lait dans une mamelle saine varie fortement selon l'espèce animale, les chèvres ayant une forte proportion de cellules épithéliales mammaires (CEM) en relation avec la production apocrine de lait.

En cas d'infection, un afflux massif de neutrophiles du sang vers la mamelle entraîne une élévation importante de la concentration de cellules somatiques (CCS) dans le lait, chez la vache (Serieys 1985b), la brebis (Morgante et al. 1996) et la chèvre (de Clermont 1992). La CCS du lait peut être considérée comme une estimation de la concentration en neutrophiles du lait et donc comme une caractérisation de l'état inflammatoire de la mamelle (Bergonier et al. 2003; Schukken et al. 2009).

L'intensité de la réponse cellulaire est très variable selon l'agent pathogène impliqué et les capacités de défense de l'animal. En général, les infections provoquées par des agents pathogènes majeurs s'accompagnent d'une élévation des CCS plus importante que celles provoquées par des agents pathogènes mineurs (Serieys 1985b; Schepers et al. 1997). En cas de mammites cliniques, les CCS peuvent atteindre plusieurs millions de cellules par ml (Pyorala et Pyorala 1997). Outre l'amplitude, la durée et l'évolution de la réaction cellulaire peuvent varier considérablement (Harmon 1994) en fonction du pathogène (Riollet et al. 1999) et de l'issue de l'infection (persistance, auto-guérison, guérison après traitement ...). Ainsi, l'utilisation de valeurs ponctuelles de CCS pour distinguer de façon instantanée les animaux sains des animaux infectés apparaît peu adéquate (Reneau 1986; Bergonier et al. 1997). Il est plus pertinent d'observer les CCS sur une période plus longue. En pratique, en France, des règles de décision ont été proposées pour les trois espèces. Elles reposent sur des CCS mesurés mensuellement sur l'ensemble de la lactation (Tableau 2). Elles constituent un bon outil de dépistage des IMI chroniques et durables (Bergonier et al. 1997). Cependant, elles ne permettent pas toujours de détecter les faibles infections ou les mammites cliniques aiguës de courte durée.

Outre le statut infectieux de la mamelle, d'autres paramètres peuvent influencer les CCS chez un animal sain. Les CCS sont généralement plus élevés en début et en fin de lactation (Serieys 1985b; de Clermont 1992; Lagriffoul et al. 1996), de part des modifications

physiologiques en début de lactation et une dilution dans un volume de lait plus petit en fin de lactation. Elles sont plus faibles lors de la traite du matin que lors de celle du soir, car l'intervalle entre deux traites est plus long, ainsi les cellules sont diluées dans un volume de lait plus important (Reneau 1986; Harmon 1994; Bergonier et al. 1994; de Clermont 1992). Néanmoins, le statut bactériologique reste le facteur prépondérant de variation des CCS (Harmon 1994; Bergonier et al. 1997).

Les CCS sont souvent mesurées à partir d'un échantillon de lait provenant de l'ensemble des mamelles de l'animal. Elles reflètent alors la moyenne des CCS de chaque mamelle, ce qui diminue le pouvoir prédictif de la mesure.

Les CCS sont souvent mesurées par une méthode directe automatique fluoro-opto-électronique, basée sur le comptage des noyaux cellulaires dont l'ADN est rendu phosphorescent (appareil Fossomatic, société Foss Electric). Le calibrage des appareils garantit une fiabilité des résultats de CCS dans une gamme allant de 0 à 1 800 000 cellules/ml (Lerray 1999). Le coût d'une mesure est relativement faible (entre 0,5 à 2 € / échantillon). Les CCS sont généralement mesurées dans le cadre du contrôle laitier officiel. Les échantillons de lait prélevés par les techniciens du FCEL sont transmis aux laboratoires d'analyses agréés qui réalisent les mesures et transmettent l'information aux Centres Régionaux d'Information. L'information est ensuite centralisée dans la base de données génétique nationale (CTIG, Centre de Traitement de l'Information Génétique, Jouy en Josas).

Tableau 2 : Règles de prédiction des IMI basées sur les CCS (* 1000 cellules/ml) mensuels

Espèce (Source) Règles de décision	Conclusion à l'animal
Vache (Serieys, 1985c)	
* Tous les CCS < 300	"non infecté durablement par un pathogène majeur"
* Au moins 2 CCS > 300 dont au moins 1 > 800	"infecté durablement par un pathogène majeur"
* Autres cas	aucune conclusion possible
Brebis (Bergonier et al, 1997)	
* Tous les CCS, sauf 1, < 400-500	"sain"
* Au moins 3 CCS > 800-1000	"infecté"
* Autres cas	"douteux"
Chèvre (de Crémoux, et al, 1994)	
* Au moins 2 CCS > 750	"infecté par un SCN" vs "sain"
* Au moins 2 ou 3 CCS > 1750	"infecté par un pathogène majeur"

Il existe principalement un second critère indirect de mesure des IMI. Celui-ci repose sur la **conductivité électrique** du lait (Fernando et al. 1982). Elle permet de mesurer les changements de concentration ionique du lait lors d'une inflammation. Elle a l'avantage de pouvoir être utilisée directement dans la salle de traite si l'appareil de traite est équipé de sondes. Elle permet de détecter 80% des mammites cliniques et 45% des mammites subcliniques (Norberg et al. 2004). Mais elle est extrêmement variable d'un animal à l'autre, ce qui limite son efficacité dans la détection des IMI au sein d'un troupeau (Hamann et Zecconi 1998). En revanche, la comparaison de la conductivité électrique entre les mamelles d'un même animal permet une bonne détection des compartiments atteints (Nielen et al. 1992; Jensen et Knudsen 1991; Pyorala 2003). Ainsi, chez la vache, qui possède quatre quartiers, l'utilisation de la conductivité électrique du lait est meilleure que chez les petits ruminants, qui ne possèdent que deux hémi-mamelles.

En résumé, il existe différents paramètres qui permettent de mesurer ou de prédire l'état infectieux de la mamelle. Chez les petits ruminants, les fréquences de cas cliniques étant faibles, les bactériologies coûteuses et les conductivités électriques difficiles à interpréter, le principal critère pour la prédiction des IMI est la mesure des CCS.

1.6 Conséquences biologiques

Les IMI qui sont le résultat de l'entrée de germes pathogènes dans la glande mammaire ont des répercussions sur la santé de la mamelle et celle de l'animal entier, ainsi que sur ses capacités de production.

Les processus inflammatoires accompagnés du recrutement de nombreux neutrophiles contribuent à augmenter la perméabilité vasculaire et tissulaire de la glande (Le Roux et al. 2003). Cette perméabilité accrue favorise le passage d'éléments d'origine sanguine dans le lait et entraîne des modifications profondes du fonctionnement de la glande mammaire.

L'augmentation des CCS s'accompagne d'une diminution de la concentration de lactose (Le Roux et al. 1995; Bansal et al. 2005; Ogola et al. 2007) et des transporteurs du glucose, des acides gras et de la L-carnitine (Ling et Alcorn 2010). Ces deux mécanismes conduisent à la diminution de la synthèse du lait. Ils peuvent s'expliquer par une moindre disponibilité des substrats et par la présence de bactéries sécrétant des toxines, qui ont la capacité d'altérer l'épithélium sécrétoire. La forte migration des neutrophiles du sang vers le lait semble également altérer l'épithélium sécrétoire (Paape et al. 1999). Ainsi les IMI sont responsables de la baisse de production laitière. Toutefois, chez la vache, un accroissement de la production laitière a été observé dans les quartiers sains compensant complètement ou partiellement la baisse de production du quartier infecté (Paapen et al 1999).

Les IMI sont aussi responsables d'une forte diminution de la concentration en protéines solubles (Bansal et al. 2005; Ogola et al. 2007) (Tableau 3), alors que la concentration en protéines totales augmente (Le Roux et al. 1995; Ogola et al. 2007). Cette augmentation est liée à l'augmentation de l'activité des plasmines (Le Roux et al. 1995), qui sont des protéinases impliqués dans les ruptures des caséines ; d'où la diminution de la teneur en caséine (Ogola et al. 2007).

Les IMI entraînent aussi la diminution de la concentration en matières grasses qui s'accompagne d'une augmentation de la teneur en acides gras libres (Ogola et al. 2007; Bansal et al. 2005) (Tableau 3). Cette observation peut s'expliquer à la fois par une augmentation de la quantité de lipases dans le lait (provenant des leucocytes endommagés) mais également par l'arrivée d'activateurs venant du sang (Serieys et al. 1987).

Les IMI sont responsables d'une augmentation de la quantité d'enzymes qui peuvent être impliquées dans les défenses contre l'infection (lysozymes, catalase) ou être un témoin de la destruction cellulaire (N-Acétyl-Glucosaminidase ou NAGase) (Welbeck et al. 2011).

La composition du lait en minéraux est également modifiée. Les teneurs en potassium, calcium et phosphore diminuent (Ogola et al. 2007) (Tableau 3). Pour le calcium et le phosphore, cela pourrait s'expliquer par une diminution des constituants auxquels ils sont complexés, comme la caséine. En revanche la teneur en chlore et en sodium d'origine sanguine augmente (Ogola et al. 2007) (Tableau 3). Globalement, le bouleversement des équilibres minéraux se traduit par une augmentation du pH du lait qui se rapproche de celui du plasma.

Toutefois, une étude a montré que les CCS n'affectent pas les capacités de transformation du lait en fromage, chez la chèvre (Raynal-Ljutovac et al. 2005). Des études sont en cours à la SAGA pour vérifier cette hypothèse.

Tableau 3 : Modifications physico-chimiques du lait causées par les mammites (revue de Serieys 1995)

Caractères physico-chimiques		Evolution
Lactose		↓
Matières protéiques	Taux protéique	=
	% protéines coagulables	↓
	% protéines solubles	↑
	Protéolyse par la plasmine	↑
Matières grasses	Taux butyreux	= ou ↓
	Acides gras libres	↑
Matières minérales	Potassium Calcium Phosphore	↓
	Sodium Chlore	↑
	pH	↑

2 Les mécanismes généraux de protection contre les infections mammaires chez les ruminants laitiers

Pour répondre aux IMI, les organismes ont mis en place des mécanismes qui permettent d'empêcher les bactéries de pénétrer dans la glande mammaire (défense naturelle), ou si elles sont entrées de les détecter et de les détruire (défenses immunitaires). Ces derniers mécanismes immunitaires reposent principalement sur l'immunité innée et l'immunité adaptative qui collaborent dans l'élimination des germes. Ces mécanismes sont présentés dans cette section.

2.1 Les mécanismes de défense naturelle

La mamelle est fortement soutenue par des ligaments permettant de maintenir les trayons verticaux et empêchant un abaissement du plancher de la mamelle qui risquerait d'être souillée plus fréquemment. La compartimentation des mamelles en quatre quartiers chez la vache et deux héli-mamelles chez les petits ruminants permet de limiter la diffusion de l'infection d'un compartiment à l'autre.

Le principal point d'entrée des agents pathogènes dans la mamelle est le **trayon** et principalement son apex, où débouche le **canal** qui conduit à la citerne mammaire. Lors de la traite, le canal s'ouvre et peut atteindre 2mm de diamètre, puis il se referme grâce au **sphincter** musculaire dans les deux heures suivant la traite. Ainsi, le sphincter empêche la pénétration des bactéries. Hors lactation, un bouchon de kératine se forme à l'extrémité inférieure du canal du trayon et permet une étanchéité totale au bout d'une semaine environ chez la vache laitière (Paulrud 2005). La **peau** qui recouvre le trayon est aussi un point critique de pénétration des agents pathogènes. Elle est glabre et dépourvue de glandes sudoripares, sébacées ou muqueuses, ce qui la rend très sensible aux conditions extérieures

(température et hygrométrie). Une diminution de son élasticité entraîne souvent des lésions de l'épiderme lors de la traite, qui sont facilement colonisables par des germes.

Toutefois la structure du trayon lui confère un rôle majeur de barrière physique et chimique (Capuco et al. 1992). La partie interne du canal du trayon est tapissée de cellules squameuses qui forment un épithélium stratifié recouvert de kératine (Paulrud 2005). La kératine par sa structure retient les bactéries et empêche leur migration vers les structures plus profondes. Elle permet l'expulsion des bactéries au moment de la traite et elle contient des substances bactériostatiques (acides gras et protéines) (Craven et Williams 1985). A l'extrémité intérieure du canal se trouve la rosette de Fürstenberg qui forme un repli muqueux qui sert de point d'entrée et d'activation des leucocytes. Ainsi, le canal du trayon est un site clé, pour limiter la pénétration des agents pathogènes dans la glande mammaire.

2.2 Les mécanismes de défense immunitaire

Ce paragraphe décrit les mécanismes de défense qui sont mis en place lorsque les mécanismes de défense naturelle ont été insuffisants et qu'un germe a franchi la barrière du canal du trayon.

En général, les défenses immunitaires sont divisées en deux catégories : l'immunité innée et l'immunité acquise. L'immunité innée (qualifiée aussi de réponse non-spécifique) est le mécanisme de défense prédominant dans la glande mammaire infectée. L'immunité adaptative passe par une reconnaissance des antigènes spécifiques d'un pathogène.

Dans la glande mammaire qui est soumise à de nombreux stress, les effecteurs de l'immunité innée, de l'immunité adaptative collaborent pour fournir une protection optimale de la glande mammaire et une élimination efficace des agents pathogènes responsables de mammites.

2.2.1 Immunité innée de la glande mammaire

L'immunité innée de la glande mammaire est composée à la fois d'éléments cellulaires et humoraux qui reconnaissent spécifiquement des molécules d'agents pathogènes. Le système immunitaire inné a trois effecteurs : le **système du complément**, les **cellules phagocytaires** et les **peptides antimicrobiens** (Rooijackers et al. 2005). Le système immunitaire inné reconnaît les antigènes étrangers, tels que les PAMP (*pathogen-associated molecular patterns, PAMP*) qui sont des molécules conservées entre espèces de microorganismes (Janeway et Medzhitov 2002). Cette reconnaissance passe par les **récepteurs PRR** (*pattern recognition receptors*), tels que les récepteurs Toll-like (*toll-like receptor, TLR*) (Vivier et Malissen 2005), ceux du complément (CLR) et d'autres récepteurs. L'immunité cellulaire innée s'articule autour des cellules phagocytaires (neutrophiles et macrophages) et des cellules naturelles tueuses (*natural killer NK*) alors que l'immunité innée humorale repose sur le système du complément, les défensines, les peptides antimicrobiens et diverses substances aux propriétés bactéricides ou bactériostatiques tels que l'oxyde nitrique, la lactoferrine, le lysozyme, la lactoperoxydase et la xanthine oxydase (Goldammer et al. 2004; Rainard et Riollet 2006; Kadowaki et al. 2001).

2.2.1.1 Immunité cellulaire

Les macrophages, les polynucléaires neutrophiles (PNN) et les cellules T représentent les populations cellulaires de la glande mammaire saine. Ces cellules sont aussi appelées des cellules somatiques par opposition aux germes exogènes. Il existe une corrélation indirecte entre les concentrations de cellules somatiques dans le lait et le niveau de protection contre la plupart des agents pathogènes provoquant des mammites (Schukken et al. 1994).

2.2.1.1.1 Récepteurs Toll-Like et défenses de la glande mammaire

La détection de la présence de bactéries dans le canal du trayon est la toute première étape d'initiation de la réponse de la glande mammaire. La présence d'agents pathogènes peut être détectée dès que le processus d'infection commence grâce aux récepteurs du système immunitaire, comme les TLR. Ces récepteurs sont exprimés à la surface d'un grand nombre de cellules de la glande mammaire (macrophages, monocytes, cellules épithéliales et neutrophiles). Ils sont impliqués dans la reconnaissance de certains PAMP. Une fois détectés, les PAMP entraînent l'activation des gènes de défense (Goldammer et al. 2004).

Tableau 4 : Récapitulatif des PRR et de leurs principaux ligands d'après Werling et Coffey 2007; Chang et al. 2009; Jungi et al. 2010; Beutler 2004

Groupe	Membres	Ligands connus
Famille des MMR	MMR	HIV, <i>Mycobacterium tuberculosis</i> , <i>Candida albicans</i>
	DEC205	Unknown
Collectins	MBL	HIV, <i>Staphylococcus aureus</i> , <i>Streptococcus pneumoniae</i> , <i>C. albicans</i> , <i>Aspergillus fumigatus</i>
	SP-A	RSV, HSV-1, <i>S. aureus</i> , <i>S. pneumoniae</i>
	SP-D	RSV, <i>M.tuberculosis</i> , <i>C. albicans</i> , <i>A. fumigatus</i> , <i>Pseudomonas aerogonosa</i>
C-type lectin R	DC-SIGN	HIV, HCV, Dengue, <i>M. tuberculosis</i> , <i>A. fumigatus</i> , <i>C. albicans</i> ,
	Dectin-1	<i>Pneumocystis</i> , <i>C. albicans</i>
Famille des TLR	TLR1-10 TLR2 TLR4 TLR5-TLR9	Peptidoglycane et acide lipotéichoïque (LTA) des bactéries Gram + Lipopolysaccharide (LPS) des bactéries Gram - Flagelline chez <i>E. coli</i>
	LBP CD14	LPS-binding protein (LBP) LPS en présence de LBP
	récepteur à FMLP	peptides fMet-Leu
	NOD1 NO D2	molécules bactériennes avec D-glutamyl-meso-diaminopimelic acid (iE-DAP)
Complément	C3b properdin	surface cellulaire des bactéries aide au fonctionnement de C3b

Actuellement, onze TLR ont été décrits chez l'homme et la souris ; chaque TLR a la capacité de reconnaître différents composés microbiens. Chez le mouton, dix TLR ont été caractérisés jusqu'à aujourd'hui (Chang et al. 2009). Par exemple, les principaux ligands de TLR2 sont les peptidoglycanes et l'acide lipotéicoïque (LTA) des bactéries Gram (+) (Takeuchi et al. 1999; Supajatura et al. 2002), alors que TLR4 reconnaît principalement les lipopolysaccharides (LPS) des bactéries Gram (-) (Miller et al. 2005). Lors d'une mammite à *S. aureus* ou *E. coli*, les gènes codant pour TLR2 et TLR4 sont exprimés (Goldammer et al. 2004; Kadowaki et al. 2001; Burton et Erskine 2003). Leur activation est suivie d'une transduction du signal par modification de cascades de gènes notamment au travers de la voie des MAPK (*mitogen activated protein kinase*) (Baranova et al. 2008). Ceci conduit à l'activation des facteurs nucléaires NFκB (*Nuclear Factor-kappa b*) et AP-1 (*Activated Protein-1*) qui déclenche l'expression de plusieurs cytokines pro-inflammatoires et permet la communication entre cellules immunitaires. Outre les TLR, il existe d'autres récepteurs de l'immunité innée à la surface des cellules impliqués dans la reconnaissance des agents pathogènes. Ils sont brièvement décrits dans le Tableau 4.

2.2.1.1.2 Les neutrophiles

Principales caractéristiques des neutrophiles

Les PNN constituent une des premières barrières de défense cellulaire contre la colonisation de la glande mammaire par des bactéries (Paape et al. 2000). La capacité de ces cellules à réagir aux stimuli inflammatoires dépend de leur aptitude à la chimiotaxie, à la phagocytose et à la destruction des composés intracellulaires par des mécanismes dépendant de l'oxygène (Paape et al. 2002). La caractéristique principale des PNN est la présence d'un noyau plurilobé, qui permet aux PNN de s'insinuer plus rapidement entre les cellules endothéliales. En général, la différenciation des PNN dure de 10 à 14 jours (Bainton et al. 1971). Une fois mûres, les PNN peuvent être stockés quelques jours supplémentaires dans la moelle osseuse. Les PNN mûres quittent le compartiment hématopoïétique de la moelle osseuse et entrent dans le sinus vasculaire en se déplaçant dans des canaux de migration entre la couche de cellules endothéliales (Figure 1). Les PNN circulent brièvement dans le système circulatoire (demi-vie de 8,9 heures) (Carlson et Kaneko 1973). En réponse à un signal chimiotactique, les PNN quittent le sang par diapédèse entre les cellules endothéliales et pénètrent dans les tissus. Là, ils agissent comme des cellules phagocytaires pendant 1 à 2 jours. Chez l'homme, la présence de produits bactériens comme le LPS augmente la durée de vie des PNN (Lee et al. 1993).

Lors d'une mammite provoquée par *E. coli*, la déplétion de l'ensemble des PNN circulants provoque des changements dans la proportion de cellules pro génitrices de la moelle osseuse (Diez-Fraile et al. 2004). Le nombre de PNN circulants avant l'infection mammaire est crucial pour l'issue de cette infection (Heyneman et al. 1990). Plus la quantité de PNN est élevée dans le sang, plus les PNN seront recrutés précocement sur le lieu de l'infection et plus les bactéries seront éliminées rapidement avant que leur développement ne dépasse les capacités de contrôle de l'organisme.

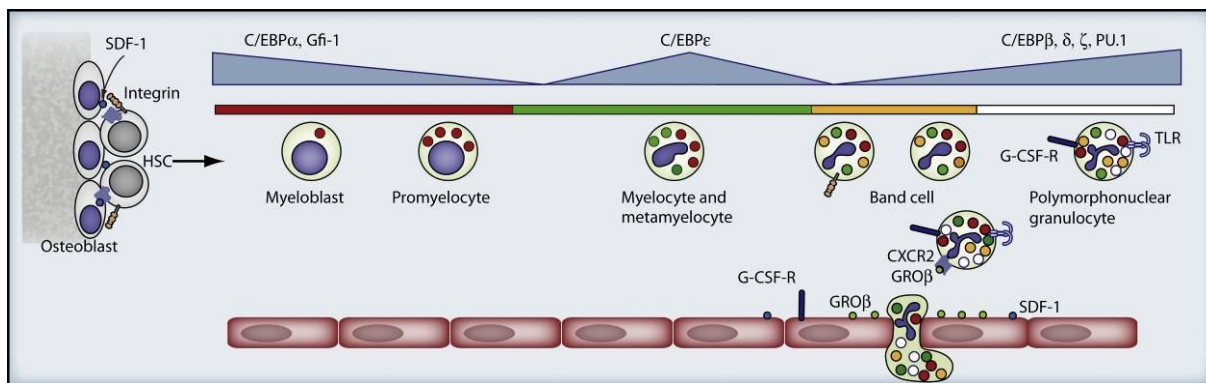


Figure 1 : Maturation des granulocytes (Borregaard 2010)

La migration des neutrophiles du sang vers le tissu mammaire

Les PNN sont les premières cellules recrutées dans le lait lorsque des agents pathogènes pénètrent dans la glande mammaire et ils représentent le type cellulaire dominant dans le lait en cas d'infection (Paape et al. 2000; Sordillo et Streicher 2002). Les PNN possèdent plusieurs récepteurs à leur surface, qui leur permettent de reconnaître des antigènes. La reconnaissance d'un antigène déclenche des processus de migration dirigée depuis le sang vers le compartiment infecté (**diapédèse**). Tout d'abord, les L-sélectines des PNN sanguins reconnaissent leurs ligands situés à la surface des cellules endothéliales (Kishimoto et al. 1989). Les PNN ralentissent et, roulent le long de l'endothélium. Puis, les intégrines $\beta 2$ des PNN (CD11a, b, et c et CD18) adhèrent aux cellules endothéliales. Les lobes du noyau sont réorganisés dans les cellules et les PNN se fauillent entre les cellules endothéliales, puis les cellules épithéliales mammaires (Figure 2). Les PNN migrent en réponse à un gradient de concentration de molécules chimio-attractantes, dont les composés du complément C3a et C5a, certaines cytokines (IL-1, IL-2) et des chimiokines (principalement CXCL8) (Taub et al. 1993). Le recrutement des PNN dépend de l'intensité et de la rapidité de diffusion de l'agent pathogène.

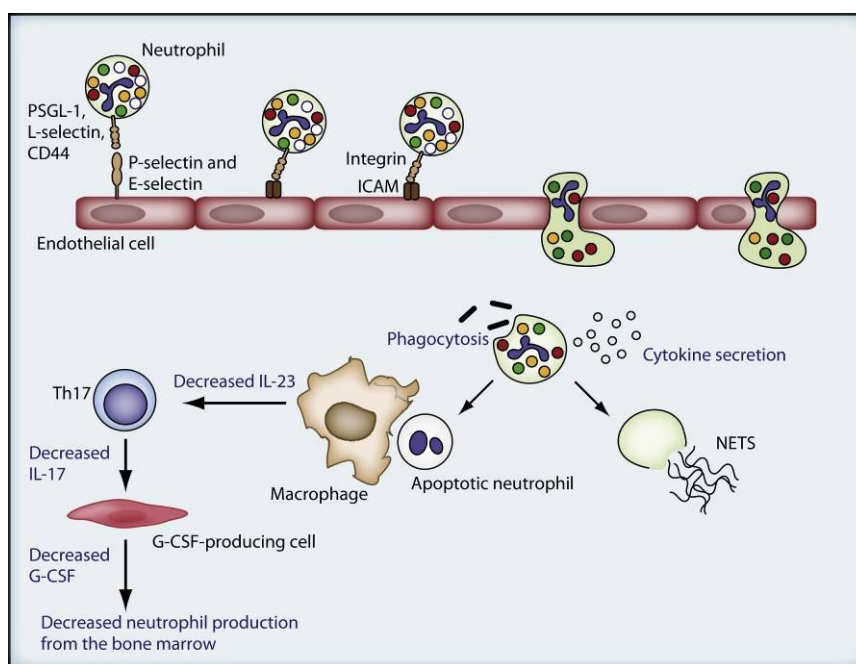


Figure 2 : Diapédèse des neutrophiles à travers la couche de cellules endothéliales (Borregaard 2010)

Dans le cas d'infection à *E. coli*, quelques heures après l'entrée des bactéries, de grandes quantités de cytokines pro-inflammatoires (IL-1 α , CXCL8) sont détectables dans le lait. Elles sont suivies par le recrutement de nombreux PNN dans les premières heures de l'infection (entre 3 et 12 heures après son début) ; le retour à des niveaux normaux se fait dans les 24 heures suivantes (Riollet et al. 2000, Shuster et al. 1997). Lors d'une infection subclinique à *S. aureus*, le recrutement de PNN est moins intense et retardé de 24 à 48 heures après le début de l'infection. Mais, si l'intensité de l'infection est plus grande, les bactéries *S. aureus* peuvent se développer dans la glande mammaire, et des neutrophiles supplémentaires peuvent être recrutés rapidement (Burton et Erskine 2003; Leitner et al. 2000).

Phagocytose et activités bactéricides des neutrophiles

Sur le lieu de l'infection, les PNN activés reconnaissent des agents pathogènes grâce à des récepteurs immunologiques (CR1, CR3, FcR) ou non-immunologiques (récepteurs aux lectines, ou glycoprotéines) et des anticorps (IgM, IgG2) situés à la surface des cellules. Puis les PNN phagocytent les bactéries (Paape et al. 2002). Certaines cytokines, comme IL-1 et TNF α peuvent favoriser la phagocytose et l'activité bactéricide des PNN.

Cependant, il a été montré *in vitro* que la diapédèse diminue les capacités de phagocytose et d'activité bactéricide des PNN lors de leur migration à travers la couche de cellules épithéliales mammaires (Smits et al. 1999). De plus, les globules gras du lait et la caséine ingérés par les PNN entraînent une diminution des granules cytoplasmiques dans le cytoplasme ce qui réduit les activités bactéricides et ralentit la formation des pseudopodes qui sont nécessaires à la phagocytose. Ainsi, les capacités fonctionnelles des neutrophiles du lait sont inférieures à celles des neutrophiles du sang.

Après ingestion des bactéries, la plupart des PNN meurent par induction de mort cellulaire programmée (apoptose). Puis des macrophages recrutés sur le site de l'infection ingèrent par phagocytose les corps apoptotiques des neutrophiles (Paape et al. 2003).

En plus de leur capacité de phagocytose, les PNN sont une source de peptides antimicrobiens (AMP, *antimicrobial peptide*), comme les lactoferrines, les bacténécines, les défensines et les cathélicidines. L'activité bactéricide des neutrophiles est permise par deux catégories de mécanismes, qui se différencient selon qu'ils dépendent ou non de l'oxygène. Lorsque certaines fractions du complément ou des immunoglobulines se fixent sur les récepteurs des PNN, ces cellules deviennent actives et génèrent des espèces réactives à l'oxygène (ROS) par l'intermédiaire de l'oxydase NADPH. Cette réponse est suivie par une augmentation de la consommation d'oxygène et de l'activité des hexoses mono-phosphates ; ces phénomènes correspondent à la chaîne respiratoire. Les PNN synthétisent alors de grandes quantités de super-oxyde (O₂⁻) et de peroxyde d'hydrogène (H₂O₂) qui interagissent entre eux et entraînent la formation d'autres radicaux libres (ROS). La plupart des agents pathogènes responsables de mammites (*S. aureus*, *E. coli*) est tuée par ces mécanismes (Paape et al. 2002).

Les activités bactéricides qui ne dépendent pas des ROS, sont liées à certaines protéines aux propriétés antibactériennes comme le complément, les défensines et les cathélicidines (Rainard et Riollet 2003; Brogden et al. 2003; Ganz 2003; Hirsch et al. 2008). Les neutrophiles peuvent aussi générer des pièges (*Neutrophil Extracellular Trap*), composés de molécules nucléaires et granuleuses qui se lient aux pathogènes et dégradent les facteurs de virulence et tuent à la fois des bactéries Gram (+) et Gram (-) (Brinkmann et al. 2004).

2.2.1.1.3 Les macrophages

Les macrophages facilitent les réponses immunitaires innée et/ou adaptative lors d'une infection bactérienne. Ces cellules sont présentes dans les glandes mammaires saines et infectées (Lee et al. 1980; Paape et al. 2002). Cependant, la proportion de macrophages dans les tissus infectés est bien inférieure à celle des neutrophiles, même si leur nombre augmente au cours de la réponse inflammatoire (Niemiłowski et al. 1988). Comme les neutrophiles, les macrophages expriment les récepteurs au fragment Fc à leur surface (Niemiłowski et al. 1988) et ont la capacité de phagocyter et tuer des agents pathogènes par l'intermédiaire de protéases ou de ROS (Miller et al. 1988). Certains anticorps opsonisants peuvent fortement augmenter l'activité de phagocytose des macrophages (Miller et al. 1988).

Les macrophages ont la capacité de sécréter des substances (prostaglandines, leucotriènes et cytokines) qui facilitent la migration et les activités bactéricides des neutrophiles, ce qui augmente fortement les processus inflammatoires locaux dans la glande mammaire (Kehrl et al. 1999; Van Kampen et Mallard 1997).

De plus, lors d'infection, les macrophages sécrètent de l'IL-12, qui peut aider au développement des cellules CD4⁺ et CD8⁺ cytotoxiques sécrétant de l'interféron γ (IFN- γ) (Riollet et al. 2000b) et conduisant à l'élimination des agents pathogènes. Bien que les macrophages ne soient pas considérés comme des cellules de défense, ils présentent des antigènes qui permettent la détection des agents pathogènes et l'initiation de la réponse inflammatoire. Lorsque la quantité d'immunoglobulines diminue, l'activité phagocytaire des macrophages décroît en parallèle (Waller 2000).

2.2.1.1.4 Les cellules dendritiques

Les cellules dendritiques se situent sous la plupart des épithéliums de surface et dans la plupart des organes solides. Elles se différencient des macrophages principalement par la présence de l'intégrine $\alpha_x\beta_2$ aussi appelée CD11c:CD18 ou récepteur 4 du complément (CR4). Tout comme les macrophages, elles ont la capacité de reconnaître des motifs moléculaires associés aux pathogènes et de les ingérer. Elles sont actives dans la capture des antigènes par phagocytose au moyen de récepteurs comme la lectine DEC205 ou le récepteur du mannose. Les antigènes captés peuvent être apprêtés et présentés sur des molécules du complexe majeur d'histocompatibilité (CMH) de classe II pour être reconnus par des cellules T CD4. D'autres antigènes extracellulaires sont prélevés de manière non spécifique par le processus de macropinocytose et sont présentés sur les molécules du CMH de classe I.

Outre les cellules dendritiques conventionnelles décrites ci-dessus, les cellules dendritiques plasmacytoïdes détectent des infections principalement avec le récepteur CXCR3 et produisent en abondance des interférons de type I et des cytokines pro-inflammatoires.

2.2.1.1.5 Les cellules naturelles tueuses

Les cellules NK sont des lymphocytes granuleux qui ont des activités cytotoxiques indépendantes du CMH mais qui utilisent des récepteurs Fc pour leur effet cytotoxique. Elles expriment à leur surface le CD16 qui est le récepteur du fragment Fc de l'IgG. Leur activité cytotoxique passe par l'exocytose de granules et la libération de cytokines. Les cellules peuvent tuer à la fois des bactéries Gram (+) et Gram (-), et peuvent participer à la protection contre les mammites (Sordillo et Streicher 2002).

2.2.1.1.6 Les cellules épithéliales mammaires

Les cellules épithéliales mammaires (CEM) sont considérées comme les premières cellules à initier le processus inflammatoire dans la glande mammaire. Elles servent non seulement de barrière physique entre le milieu extérieur et l'organisme en empêchant la pénétration des agents pathogènes, mais elles jouent aussi un rôle dans la réponse immunitaire. Les CEM après stimulation par des bactéries ont la capacité de sécréter des cytokines, des chimiokines et des peptides antimicrobiens (Lutzow et al. 2005; Griesbeck-Zilch et al. 2009). Leur rôle principal dans la glande mammaire pourrait être leur contribution au recrutement des neutrophiles et des lymphocytes.

Généralement, la reconnaissance des agents pathogènes dans la glande mammaire intervient à travers les récepteurs TLR. Les gènes codant pour TLR2 et TLR4 sont exprimés dans les CEM, même si leur expression au niveau protéique n'a pas été formellement démontrée. Ainsi, comme les CEM répondent bien à la stimulation par *S. aureus* et *E. coli*, il est probable que les CEM expriment certains récepteurs TLR.

2.2.1.2 La fraction soluble de l'immunité innée

Dans la glande mammaire, les principaux effecteurs de la réponse humorale sont les peptides antimicrobiens et les composés du système du complément.

2.2.1.2.1 Les peptides antimicrobiens

Les peptides antimicrobiens sont des petits polypeptides qui ont des effets antimicrobiens à des concentrations physiologiques dans l'organisme (Brogden et al. 2003; Ganz 2003). Ils sont présents dans différentes populations cellulaires comme les PNN, les macrophages et les cellules épithéliales. Les principaux peptides antimicrobiens chez les bovins sont les défensines, les cathélicidines et les peptides anioniques (Brogden et al. 2003).

Les défensines sont produites en grande quantité dans les cellules polynucléaires (Zasloff 2002; Goldammer et al. 2004). Chez les bovins, elles ont des activités antimicrobiennes avérées contre *S. aureus* et *E. coli* (Lutzow et al. 2005). Les défensines peuvent être exprimées dans la mamelle saine ou être induites lors d'une infection. Elles sont supposées jouer un rôle important dans la défense de la glande mammaire.

Les cathélicidines sont les principales protéines des granules des neutrophiles. Leur production est augmentée en présence d'infection (Zanetti 2005). Lors de la dé-granulation, les cathélicidines sont clivées par des protéases et libèrent un peptide C-terminal qui a des activités antimicrobiennes (Tomasinsig et Zanetti 2005).

Les lactoferrines sont produites par les cellules épithéliales et les leucocytes. Elles chélatent le fer et le rendent indisponible pour la croissance des microorganismes (Sordillo et Streicher 2002; Rainard et Riollot 2006). L'expression des lactoferrines semble fortement augmentée en présence de LPS (Schmitz et al. 2004), mais leur rôle dans la réponse immunitaire n'est pas bien connu.

2.2.1.2.2 Le système du complément

Le système du complément est impliqué surtout dans l'immunité innée avec la lyse des bactéries, l'opsonisation et l'attraction des phagocytes sur le lieu de l'infection (Rainard et

Poutrel 1995; Riollet et al. 2000b; Korhonen et al. 2000; Rainard et Riollet 2006). Le système du complément est activé par trois voies métaboliques différentes, qui conduisent toutes à l'activation de la convertase C3. La convertase est associée au complexe enzymatique qui clive C3. Ce clivage active le positionnement de C3b à la surface des microorganismes et libère C3a. C3a a un fort pouvoir chimio-attractant pour les cellules inflammatoires. C3b et ses sous-produits iC3b sont déposés à la surface des bactéries et permettent leur reconnaissance par les cellules phagocytaires (neutrophiles et macrophages). La protéine C5a du complément est également impliquée dans le recrutement des neutrophiles (Persson et al. 1993). On la retrouve dans le lait des vaches saines (Rainard et Poutrel 1995) et à plus forte concentration dans le colostrum ou lors d'infection surtout par *E. coli* ou *S. uberis*. Mais il semblerait qu'elle ne soit pas augmentée dans le cas d'infection à *S. aureus* (Shuster et al. 1997; Riollet et al. 2000b; Bannerman et al. 2004).

2.2.2 La réponse immunitaire acquise de la glande mammaire

Les principaux effecteurs de l'immunité acquise sont les anticorps, les lymphocytes B et les cellules T qui agissent ensemble sous l'action de diverses cytokines (Burton et Erskine 2003; Burvenich et al. 1994). Plusieurs types cellulaires de la glande mammaire, comme les cellules dendritiques, les cellules NK, les lymphocytes T $\gamma\delta$, les macrophages, les cellules CD4⁺ et CD8⁺ jouent un rôle important dans la réponse inflammatoire locale (Burton et Erskine 2003).

2.2.2.1 Les lymphocytes et les défenses mammaires

Les lymphocytes représentent la population cellulaire majeure de l'immunité acquise, car ils sont capables de détecter des antigènes grâce à leurs récepteurs de surface, qui déterminent leur spécificité, leur diversité, la discrimination du soi et du non-soi (Sordillo et Streicher 2002; Park et al. 1992). Ils sont divisés en deux principales catégories : les lymphocytes B et les lymphocytes T.

Les lymphocytes B produisent les immunoglobulines après stimulation par des cytokines produites par les cellules T auxiliaires. Ces cytokines (IL-2, -4, -6 et -10) induisent la prolifération et la différenciation des lymphocytes B, soit en plasmocytes qui produisent des anticorps, soit en cellules mémoires. La proportion de lymphocytes B par rapport à la population totale de lymphocytes reste à peu près constante après une infection mammaire (Riollet et al. 2000a).

2.2.2.2 Les cellules B et les immunoglobulines

Les immunoglobulines (Ig) ou anticorps sont sécrétés par les lymphocytes B et sont présents dans tous les liquides biologiques de l'organisme, à des concentrations variables. Les lymphocytes B, activés par les antigènes bactériens, prolifèrent, migrent depuis l'épithélium sécrétoire vers les muqueuses et se différencient en cellules sécrétant massivement ces anticorps. Chez les ruminants laitiers, quatre classes d'immunoglobulines ont un rôle avéré dans le contrôle d'une infection (Guidry et Miller 1986): les IgA et les IgM synthétisées localement, les IgG1 arrivant du sang par transport sélectif et les IgG2 affluant du sang de façon passive par l'adhérence aux neutrophiles. Les IgA semblent empêcher l'adhésion des bactéries aux CEM et faciliter ainsi leur évacuation au moment de la vidange (Poutrel 1983). Elles pourraient aussi contribuer à neutraliser les toxines produites par les agents pathogènes

microbiens. Il a été montré que la concentration en IgA dans le lait est corrélée à l'activité phagocytaire des neutrophiles vis-à-vis de *S. aureus* (Burton et Erskine 2003). Les IgM, IgG1 et IgG2 jouent un rôle d'opsonines (Poutrel 1983; Rainard et Poutrel 1993; Sordillo et al. 1997). En effet, ces immunoglobulines reconnaissent et se fixent aux antigènes bactériens, directement ou avec des fragments du complément. Associées aux bactéries, elles se fixent à des récepteurs spécifiques des neutrophiles et macrophages, ce qui permet d'augmenter l'efficacité avec laquelle les cellules phagocytaires ingèrent et détruisent les microorganismes infectants. Les IgG2 et IgM sont principalement impliquées dans l'élimination de *S. aureus* par phagocytose après opsonisation (Burton et Erskine 2003). La forme IgG1 est l'immunoglobuline dominante dans le lait sain et le colostrum. Elle semble être impliquée dans la prévention de la colonisation bactérienne et la neutralisation des toxines (Burton et Erskine 2003).

2.2.2.3 Le rôle des cytokines dans la glande mammaire

Les cytokines régulent et coordonnent les réponses immunitaires par stimulation de l'activité de certaines cellules qui sont impliquées dans ces voies. Elles forment un réseau complexe qui régule les réponses immunitaires cellulaires ou humorales. Les cytokines comme IL-2 et IFN γ sont produites par exemple par les lymphocytes Th1. La voie Th1 peut éliminer la plupart des infections mammaires (Burton et Erskine 2003) et particulièrement celles à *S. aureus* (Lin et al. 2009). Au contraire, IL-4, -5 et -10 sont produites par des cellules Th2 et favorisent les réponses immunitaires humorales (Shafer-Weaver et al. 1999; Riollet et al. 2000a). Il existe une troisième voie qui repose sur les cellules Th17, qui sécrètent de l'IL-17a ou IL-17f entraînant la sécrétion de cytokines pro-inflammatoires (IL-6, TNF α , IL-1 β , ...) et des chimiokines (CXCL8, CXCL1, CXCL10, ...) (Leung-Theung-Long et Guerder 2008).

Lors d'une infection mammaire, de nombreuses cytokines sont produites en grande quantité (paragraphe 2.3). Parmi elles, l'interleukine 8 (IL-8, CXCL8) est la principale cytokine responsable de la migration des neutrophiles (Kobayashi 2008). Les cytokines pro-inflammatoires induisent localement l'expression de molécules d'adhésion sur les cellules endothéliales et favorisent ainsi la chimiotaxie des neutrophiles sur le site de l'infection. Elles peuvent aussi entraîner une augmentation de l'activité phagocytaire et bactéricide des neutrophiles (Sordillo et Streicher 2002) en augmentant l'expression des récepteurs Fc à la surface cellulaire.

De plus, les facteurs de croissance ou CSF (*colony stimulating factor*), MCSF (*macrophage CSF*) et les GMCSF (*granulocyte MCSF*) sécrétés en réponse à une infection, stimulent le développement et la différenciation de certaines cellules hématopoïétiques. Ainsi, ils jouent un rôle clé dans le déclenchement de la prolifération et le recrutement des cellules sur le site de l'infection.

Les cytokines ont la capacité de réguler la réponse immunitaire et permettent ainsi une bonne protection de la glande mammaire contre les agressions des agents pathogènes avec un minimum d'effets collatéraux sur le parenchyme mammaire. Leur étude est primordiale dans la compréhension des mécanismes cellulaires de réponse aux infections.

2.3 Variation du transcriptome lors d'infections mammaires

La puce à ADN (appelée microarray) permet de mesurer l'expression de centaines, voire de dizaines de milliers de gènes, simultanément. Ainsi, c'est un outil de choix pour comprendre les réponses cellulaires aux infections. En particulier, pour les IMI, plusieurs

études ont utilisé cette technologie sur différents types cellulaires : cellules somatiques du lait (CSL) prélevées *in vivo*, cellules épithéliales mammaires (CEM), macrophages ainsi que sur d'autres tissus (foie, tissu adipeux). Dans ce paragraphe, je me concentrerai sur les analyses du tissu mammaire.

Brièvement, l'infection de CSL soit par des bactéries Gram (+), soit par des bactéries Gram (-) conduit à l'augmentation de cytokines pro-inflammatoires, telles que *il1b*, *il8*, *il12*, *ifn γ* et *tnfa* (Bannerman 2009), de marqueurs de l'inflammation, tels que la serum amyloïde et le *lipopolysaccharide* (LPS) *binding protein* (LBP) (Bannerman et al. 2004) et aussi de sous-unités du complément principalement *c3* et *c5a* impliqués dans le recrutement des neutrophiles et la stimulation de leurs activités bactéricides (Bannerman et al. 2004; Rainard et Riollot 2006). Au moment de la résolution de l'infection, des cytokines régulatrices de l'inflammation sont activées, telles que *il10*, mais elles sont absentes lors d'une infection à *S. aureus* (Bannerman et al. 2004). La comparaison de CSL infectées et de CSL saines a montré l'importance des cytokines, des chimiokines et de leurs récepteurs. Par exemple chez la chèvre, il y a plus d'*il1a*, de *cxcl6* et de *il2 γ* dans les héli-mamelles infectées que dans les saines (Pisoni et al. 2010). Chez la vache, *il6*, *il8*, *il12*, *gmcsf* et *tnfa* sont plus exprimés dans les quartiers infectés (Alluwaimi et al. 2003; Lee et al. 2006), alors que *il4*, *il10* et *ifn γ* sont plus exprimés dans les quartiers sains (Peli et al. 2003). Lors d'infection à *S. aureus*, des cytokines (*ccr8*, *rantes*, *il17* et *mcsf*), des molécules impliquées dans la présentation des antigènes telles que *mhcI* et *tap2* et des molécules de reconnaissance telles que *cd14* et *icam3* sont activées (Tao et Mallard 2007), alors que les expressions d'*il2* et d'*ifn γ* diminuent au cours de l'infection (Alluwaimi et al. 2003).

En ce qui concerne, les réponses des CEM *in vitro* après stimulation par *E. coli*, des augmentations des facteurs immunitaires, tels que *ccl5*, *il6*, *il1*, *nos2*, *ccl20*, *il8*, *ccl2*, *cxcl5* et *cxcl2* sont observées après 1h de stimulation (Gunther et al. 2009). Après stimulation par *Streptococcus uberis* l'expression de cytokines pro-inflammatoires est augmentée, alors que des gènes relatifs à l'immunité ne sont pas modifiés (*c3*, *il1 β* , *il6*, *il8*, *saa3*, *s100a&é*, *tlr2*) (Swanson et al. 2009). Une stimulation de CEM par *S. aureus* entraîne aussi l'augmentation de cytokines pro-inflammatoires, telles que *cxcl1*, *cxcl2*, *ifn γ* , *il6*, *il8*, *il1 β* et la leptine, ainsi que de *s100a12* et *ptx3* dont l'expression a été vérifiée à un niveau protéique (Lutzow et al. 2008). Ainsi, les CEM ont la capacité d'initier la réponse immunitaire *in vivo*.

Toutefois, au sein de la glande mammaire, il existe de grandes différences des réponses cellulaires en fonction de la localisation des cellules et de la cinétique de pénétration des agents pathogènes. Ainsi, la réponse du trayon est plus rapide et plus forte que celle de la glande mammaire et la rosette de Furstenberg sécrète plus de *nos2* qui est impliqué dans la production de molécules bactéricides (Rinaldi et al. 2010). La réponse dépend aussi des populations de cellules résidentes et des neutrophiles recrutés qui ne le sont pas de façon homogène dans la glande mammaire (Rinaldi et al. 2010).

Récemment, il a été montré que des macrophages cultivés *in vitro* avec du LPS exprimaient *tlr4*, *tlr2* et *cd14*, mais que cette activation ne passait pas par la voie de MYD88 et que les cytokines pro-inflammatoires *tnfa*, *il6*, *il8* et *il1b* étaient exprimées, ainsi que la cytokine anti-inflammatoire *il10* et le gène *nos2* (Taraktsoglou et al. 2011). Mais il n'y a pas d'information disponible sur la réponse des macrophages vis-à-vis de *S. aureus*.

Ainsi, les réponses cellulaires après une infection dépendent de nombreux facteurs, dont le type cellulaire et la bactérie. La réponse à *E. coli* se traduit par des sécrétions de cytokines plus rapides et plus fortes que celle à *S. aureus* (Bannerman 2009). Le type de stimulation (bactérie vivante, bactérie inactivée, ou PAMP) et la durée de la stimulation

influencent aussi les résultats. Cependant, les capacités de réponse de l'hôte semblent être un facteur déterminant pour les réponses cellulaires aux infections (Burton et Erskine 2003).

En 2008 au début de ma thèse, aucune étude prenant en compte la variabilité de l'hôte en réponse aux IMI n'avait été réalisée. C'est pourquoi, il était intéressant d'étudier la réponse aux infections en fonction de la sensibilité / résistance de l'hôte. Depuis l'équipe de Schwerin a développé deux modèles bovins ayant des capacités de réponse aux IMI différentes. Dans le premier les vaches ont été sélectionnées sur des SCS extrêmes et dans le deuxième sur un QTL associé au SCS et positionné sur le chromosome BTA18 (Kühn et al. 2003 ; Kuhn et al. 2008). Les deux modèles et le nôtre permettent d'améliorer la compréhension des interactions hôtes-agents pathogènes.

3 Lutte contre les mammites : des méthodes curatives et préventives

Les défenses intrinsèques de la mamelle et les défenses immunitaires ne sont pas suffisantes pour éviter les IMI. Des actions peuvent alors être mises en place pour soigner les animaux atteints et pour prévenir les infections. Dans ce paragraphe, je vais tout d'abord présenter les méthodes de lutte classique contre les IMI : méthode curative (traitements antibiotiques) et traitements préventifs (suivi sanitaire du troupeau et vaccination). Dans un deuxième temps j'aborderai les efforts collectifs qui sont mis en place pour faire diminuer l'incidence des mammites : l'amélioration génétique. Je commencerai par décrire les paramètres génétiques et présenter comment ils sont utilisés en sélection, puis je présenterai les outils de la génétique moléculaire ainsi que leur utilisation en sélection.

3.1 Méthodes de lutte classique

3.1.1 Des traitements curatifs : les antibiotiques

En cas d'IMI, les traitements vétérinaires reposent essentiellement sur l'utilisation d'antibiotiques. Selon la gravité des symptômes ils sont soit appliqués par voie intramammaire localement, soit par voie générale. Lorsque les symptômes cliniques sont très prononcés, les deux voies peuvent être associées. Les antibiotiques permettent de réduire significativement l'incidence des mammites. Mais ils ne sont pas efficaces lorsque les bactéries ont pénétré à l'intérieur des neutrophiles ou des CEM (Sutra et Poutrel 1994; Aungier et Austin 1987).

Le respect de la posologie est un critère déterminant pour l'efficacité du traitement. Une diminution de la durée de traitement augmente les risques de rechute des vaches (Barkema et al. 2006). Ces risques de rechute sont souvent liés au fait que les bactéries étaient toujours présentes dans la mamelle et qu'elles se multiplient à nouveau après l'arrêt du traitement antibiotique. De nos jours, l'utilisation excessive d'antibiotiques chez l'homme a conduit à l'apparition de forme de résistance de *S. aureus* à la méthilicine (MRSA). Récemment, des souches de MRSA ont été identifiées dans des troupeaux laitiers (Spohr et al. 2011). Pour diminuer l'impact de ces nouvelles souches, il est recommandé de diminuer l'utilisation d'antibiotiques (Teuber 2001; Phillips et al. 2004). Auparavant, un traitement antibiotique des vaches était recommandé systématiquement lors du tarissement. Aujourd'hui seules les vaches infectées reçoivent un traitement antibiotique. Le reste du troupeau peut recevoir un traitement préventif, sans antibiotique ou rien du tout.

Les techniques d'identification des agents pathogènes (diagnostic bactériologique) permettent d'adapter le traitement à l'agresseur. Une bonne connaissance des réponses immunitaires chez les ruminants peut aussi permettre de développer de nouvelles alternatives pour lutter contre les mammites.

3.1.2 Des méthodes préventives

Certaines pratiques préventives peuvent aider à maintenir un bon état sanitaire du troupeau.

3.1.2.1 Hygiène et gestion du troupeau

Une thérapie du troupeau hors lactation, accompagnée de bonnes pratiques hygiéniques, notamment au moment de la traite, sont des outils efficaces pour contrôler les mammites à *Staphylococcus* (Hutton et al. 1990). Pour diminuer l'incidence des mammites il est conseillé : i) de laver les trayons avant la traite, avec une serviette différente pour chaque animal, ii) de tremper les trayons dans une solution désinfectante après la traite, iii) d'utiliser des gants et iv) de désinfecter la salle de traite (Sutra et Poutrel 1994). Il est aussi important de tester et d'entretenir régulièrement la machine à traire pour éviter des problèmes de sur-traite qui peuvent entraîner des lésions du trayon (Hillerton et al. 2002).

Le fait de nourrir les vaches juste après la traite permet au trayon d'avoir le temps de se refermer avant que les vaches ne se couchent et que le trayon soit en contact avec la litière souillée (Bartlett et al. 1992). Récemment, il a été montré que le respect des règles ci-dessus permettait de diminuer les CCS (Barkema et al. 1998; Plozza et al. 2011). De plus, le traitement et l'isolement si nécessaire des animaux atteints permettent de limiter les contaminations entre animaux. L'élimination des animaux avec des mammites chroniques ou récurrentes permet aussi de mieux gérer l'état sanitaire global du troupeau (Neave et al. 1969).

L'obturation des trayons lors du tarissement pour limiter l'entrée des agents pathogènes dans la mamelle permet de limiter les IMI.

L'ensemble de ces pratiques permet de diminuer l'exposition des mamelles à des agents infectieux et de limiter la transmission des IMI entre différents animaux. Cependant, elles ne permettent pas d'éradiquer les mammites.

3.1.2.2 Vaccination

Depuis quelques années, de nombreux efforts ont été réalisés pour développer des vaccins efficaces contre les agents responsables des mammites. Malgré quelques essais (Middleton et al. 2006; Leitner et al. 2003; Middleton et al. 2009; Lee et al. 2005; O'Brien et al. 2001), aucune immunité prometteuse et à longue durée n'a été obtenue contre les infections à *S. aureus*. Pour la plupart des vaccins les essais terrain n'ont pas été concluants à l'exception de celui de l'équipe de Leitner (2003). Ils ont montré qu'après vaccination, les CCS de vaches en première lactation étaient plus faibles (moins 42%) et que les vaches produisaient 0,5kg de lait de plus par jour, même si la fréquence des cas cliniques dans les deux groupes de vaches n'était pas différente (1,3% contre 2,7%) (Leitner et al. 2003). Quelques essais ont aussi été menés chez la brebis, mais les résultats ont montré des efficacités variées (Watson 1988; Amorena et al. 1994).

L'efficacité des vaccins dépend des molécules de *S. aureus* incorporées : i) cellules vivantes, atténuées ou tuées, ii) toxines (hémolysines α et β), iii) PAMP (Calzolari et al. 1997; Gilbert et al. 1994; Mamo et al. 1994; Watson et al. 1996; Pankey et al. 1985), mais aussi des adjuvants et de la voie d'administration du vaccin (O'Brien et al. 2001; Tollersrud et al. 2002; Lee et al. 2005). L'efficacité des vaccins peut être limitée : i) par l'accessibilité de certains PAMP si présence de capsules autour des bactéries, ii) par l'évolution rapide des souches bactériennes et de leurs PAMP. Les mammites étant causées principalement par des bactéries d'espèces différentes, des études de vaccins à large spectre sont en cours.

En France, le vaccin STARTVAC des laboratoires HIPRA dirigé contre *S. aureus* et *E. coli* est utilisé chez les vaches et les brebis laitières. Il permet de diminuer les mammites à *E. coli*, *S. aureus* et à staphylocoques à coagulase négative, de réduire la gravité des symptômes, d'accélérer la guérison des mammites cliniques et de diminuer la contagion du milieu extérieur en diminuant le nombre de vaches infectées.

Toutefois, même si certains vaccins ont une efficacité reconnue pour diminuer l'incidence des mammites cliniques et subcliniques, ils ne permettent pas actuellement d'éradiquer les IMI. Une autre approche préventive consiste à favoriser les défenses immunitaires des animaux et à sélectionner des animaux plus résistants.

3.2 L'amélioration génétique

Pour pallier les problèmes d'efficacité des vaccins et limiter l'impact des IMI sans recourir à l'utilisation systématique des traitements antibiotiques, une alternative consiste à améliorer les capacités intrinsèques de l'hôte à résister aux IMI. Cette stratégie présente l'avantage d'être durable puisque les gains de résistance se transmettent de générations en générations.

Dans cette section, je commencerai par définir la résistance aux IMI. Puis je donnerai des définitions et des valeurs des principaux paramètres génétiques des critères liés aux IMI. Ensuite, j'aborderai brièvement l'utilisation de ces critères en sélection. Enfin, je présenterai les bases de la génétique moléculaire et ses apports en sélection par rapport à une sélection sur index.

3.2.1 Qu'est-ce que la résistance et comment la mesurer ?

Le caractère de résistance correspond à la capacité des animaux i) à éviter l'entrée des agents pathogènes dans l'organisme, ii) à augmenter les réponses immunitaires capables de limiter le développement des agents pathogènes dans l'organisme, iii) à mieux les éliminer et iv) à contrôler les effets des agents pathogènes (par exemple, la dégradation des tissus). En revanche, un animal tolérant aux mammites peut héberger un agent pathogène qui peut croître, l'hôte ne présentant pas de dégâts liés à la présence de l'agent pathogène (Medzhitov 2009) et un animal résilient est un animal tolérant dont le niveau de production n'est pas diminué en présence des agents pathogènes. Les animaux tolérants et résilients peuvent donc représenter des sources de contamination pour les autres animaux d'un troupeau. Il est donc préférable de les réformer.

Concernant les mammites, les animaux tolérants ou résilients sont des animaux ne présentant pas de signes cliniques, mais dont la bactériologie du lait est positive, alors que les animaux complètement résistants ont une bactériologie du lait négative. Le compartiment infecté (tissu mammaire) étant directement en contact avec le lait, la présence d'agents

pathogènes même s'ils n'entraînent pas des signes cliniques, provoquent une contamination du lait qui est problématique pour la consommation humaine. Ainsi, les objectifs de l'amélioration génétique se focalisent sur des animaux résistants et non pas tolérants ou résilients.

Les gènes ou les mécanismes mis en place chez les animaux résistants n'étant pas connus, un moyen d'approcher la résistance est de distinguer les animaux malades des animaux sains. En effet, les mammites étant des maladies infectieuses multifactorielles et les élevages laitiers étant de plus en plus intensifs, la pression d'infection est généralisée, ce qui signifie que tous les animaux peuvent être en contact avec des agents pathogènes responsables de mammites. Ainsi, à l'échelle d'une population, les animaux qui ne sont jamais atteints peuvent être considérés comme des animaux résistants. Dans le cadre des mammites, pour discriminer les animaux atteints des animaux sains, le statut bactériologique de la glande mammaire est mesuré à un moment donné ou pendant une période de référence.

Sa mesure repose sur les différents critères qui ont été développés dans le paragraphe 1.4 avec des mesures qui correspondent au diagnostic exact des infections (diagnostic bactériologique du lait et détection des cas cliniques) et des mesures qui permettent prédire le statut infectieux de la mamelle selon l'intensité de la réponse inflammatoire (conductivité du lait et CCS). Dans ce paragraphe, je ne développerai pas la conductivité électrique du lait. Pour que le critère mesuré puisse être utilisé en sélection, il faut qu'il soit pertinent et donc bien corrélé avec le phénotype recherché, qu'il soit mesurable facilement et à grande échelle, qu'il soit peu coûteux et qu'il soit héritable, c'est-à-dire qu'il existe une variabilité génétique.

3.2.2 Les paramètres génétiques

Les différents critères de mesure de la résistance aux mammites sont considérés comme des caractères polygéniques, c'est-à-dire qu'il existe un grand nombre de gènes qui ont un petit effet sur ces caractères et qui expliquent l'état de résistance d'un animal. En génétique quantitative, l'estimation de paramètres génétiques permet de quantifier, sans connaissance *a priori* de ces gènes, l'origine génétique de la résistance.

La connaissance des paramètres génétiques de ces caractères est indispensable pour choisir les objectifs et critères de sélection. Elle détermine le progrès génétique attendu. Je vais commencer par donner les définitions des principaux paramètres génétiques, puis je donnerai des informations sur les valeurs de ces paramètres pour les caractères de résistance aux mammites.

Définitions

Variance génétique additive (V_g) : les variations d'un phénotype sont toujours dues à des facteurs environnementaux (alimentation, climat, ...) et à des différences entre les génotypes individuels, transmissibles à la descendance, celles-ci correspondent à la variabilité génétique.

$$\sigma_p^2 = \sigma_g^2 + \sigma_i^2 + \sigma_e^2$$

$$\sigma_g^2 = \sigma_a^2 + \sigma_d^2$$

où σ_p^2 est la variance phénotypique, elle correspond à la variance totale,

σ_g^2 la variance génétique,

σ_a^2 la variance génétique additive

σ_d^2 la variance de dominance

σ_i^2 la variance d'épistasie

σ_e^2 la variance environnementale.

Héritabilité (h^2) : l'héritabilité est une donnée statistique évaluant la part de la variabilité génétique dans la probabilité d'apparition d'un caractère phénotypique donné au sein d'une population donnée. On la calcule de la façon suivante :

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Corrélation génétique (r_g) : proportion de variance que deux caractères partagent due à des causes génétiques. Pour deux caractères a et b, on la calcule avec la formule suivante :

$$r_g = \frac{\text{cov ar}(a,b)}{\text{var}(a)\text{var}(b)}$$

Bactériologie du lait

L'héritabilité du diagnostic bactériologique est estimée à 0,02 et 0,04 (Weller et al. 1992). Ainsi, il existe une variabilité génétique liée à ce caractère, mais elle est très faible par rapport à la variance environnementale.

Mammites cliniques

Il en est de même pour les mammites cliniques (MC) qui ont aussi une héritabilité très faible entre 0,02 et 0,04 chez les vaches (Mrode et Swanson 1996; Detilleux 2002; Rupp et Boichard 2003; Rupp et Foucras 2010).

Chez les ovins laitiers et les caprins, aucune donnée de paramètres génétiques n'est disponible pour les mammites cliniques aiguës ou subaiguës car leur fréquence est faible dans la population (5%).

Les concentrations de cellules somatiques

Les paramètres génétiques ne sont pas calculés directement sur les CCS, mais sur les scores de cellules somatiques (SCS) qui sont calculés à partir d'une transformation logarithmique des CCS (Ali et Shook 1980) permettant de normaliser la distribution des données :

$$SCS = \log_2 \left(\frac{CCS}{100000} \right) + 3$$

Les SCS prennent généralement des valeurs continues entre 0 et 9. La valeur 3 correspond à un CCS de 100 000 cellules/ml et une augmentation de 1 unité du SCS correspond à un doublement des CCS.

Les estimations de l'héritabilité des SCS sont de 0,15 en moyenne chez les bovins laitiers (Mrode et Swanson 1996; Detilleux 2002; Rupp et Boichard 2003), entre 0,11 et 0,14 chez les ovins laitiers (revue Rupp et al. 2010) et entre 0,20 et 0,24 chez les caprins laitiers (Rupp et al. 2011).

La corrélation génétique entre les SCS et l'infection bactérienne est proche de 1 chez la vache (Weller et al. 1992), ainsi les CCS et les mammites subcliniques correspondent presque au même caractère. Celle entre les SCS et les mammites cliniques est de 0,70 en moyenne avec des valeurs allant de 0,30 à 0,80 selon les auteurs (Mrode et Swanson 1996; Rupp et Boichard 2003; Carlén et al. 2004; Koivula et al. 2005; Bloemhof et al. 2009; Bonaiti et al. 2005) ou 0,60 (Heringstad et al. 2000). Ces valeurs relativement élevées suggèrent des mécanismes communs qui conduisent à une meilleure résistance aux IMI modérées à sévères (avec des SCS élevés) et aux MC, malgré le fait que différents pathogènes puissent être impliqués dans ces deux types d'IMI.

Par ailleurs, des associations génétiques favorables modérées (de 0,20 à 0,50) sont généralement rapportées entre SCS ou MC d'une part et des caractères de morphologie mammaire d'autre part (Barillet et al. 2007; Bonaiti et al. 2005). Ceci indique que l'anatomie de la mamelle participe à la résistance de l'animal aux IMI.

D'après la littérature, il existe un antagonisme génétique entre la résistance aux mammites (SCS et MC) et les caractères de production, ce qui a probablement conduit à la dégradation de la santé de la mamelle en conséquence de la sélection sur les caractères de production. La corrélation génétique entre SCS et rendement en lait se situe généralement entre 0,10 et 0,20 chez la vache (Mrode et Swanson 1996; Rupp et Boichard 2003; Carlén et al. 2004; Koivula et al. 2005; Rupp et Foucras 2010). Les valeurs chez les ovins sont très différentes d'une étude à l'autre, allant de valeurs antagonistes (Rupp et al. 2003; Riggio et al. 2007) à favorables (el-Saied et al. 1999; Othmane et al. 2002; Serrano et al. 2003; Legarra et Ugarte 2005), mais toujours modérées. Chez les chèvres, ces valeurs varient de -0,13 à 0,12 (Rupp et al. 2011). Il existe aussi des corrélations entre la résistance aux mammites et la conformation de la mamelle chez les vaches (Mrode et Swanson 1996; Rupp et Boichard 2003), les brebis (Legarra et Ugarte 2005; Allain et al. 2010) et la chèvre (Rupp et al. 2011). Des mamelles plus hautes et plus étroites, mieux attachées, sont associées à des SCS plus faibles. La conformation de la mamelle est donc un prédicteur de la santé de la mamelle, sans pour autant que l'on puisse clairement distinguer cause et conséquence.

La réponse immunitaire

Outre les données CCS et mammites cliniques, quelques auteurs se sont concentrés sur des caractères en relation avec les mécanismes de défense de l'hôte (revue Rupp et al. 2010). En particulier, les aspects génétiques du fonctionnement des neutrophiles ont été étudiés (Sordillo et al. 1997; Rainard et Riollot 2006). L'existence d'un effet mâle a été mise en évidence pour la phagocytose *in vitro* des neutrophiles du sang (Lostrie-Trussar et al. 1984; Kehrlé et al. 1991; Macdonald et al. 1994). La migration (0,2 à 0,5) et la phagocytose (0,3 à 0,7) des neutrophiles et l'activité du complément (0,4 à 0,5) ont des héritabilités modérées (Detilleux et al. 1994). Des animaux avec de faibles fréquences d'IMI ont tendance à avoir un meilleur fonctionnement de leurs neutrophiles (Kelm et al. 1997).

D'autres études ont montré que la réponse immunitaire par les anticorps a une héritabilité modérée à forte, de 0,32 à 0,64 pour les anticorps du sérum après vaccination, et de 0,13 à 0,88 pour les antigènes de *E. coli* pour des vaches au moment de la parturition (Wagter et al. 2000).

L'ensemble de ces mécanismes pourraient donc jouer un rôle important dans le déterminisme génétique de la résistance aux mammites, mais leur relation génétique avec la résistance reste peu explorée.

3.2.3 Génétique quantitative et sélection

Pour un caractère donné, si les paramètres génétiques établis ont des valeurs d'index et de précision suffisantes, il pourra être utilisé en sélection. Un caractère pertinent doit être héritable, bien corrélé avec le caractère sur lequel on veut agir et facile à mesurer à grande échelle et à faible coût.

Les objectifs des programmes de sélection se basent sur une diminution de la fréquence des mammites cliniques et des mammites subcliniques. A cause de leurs coûts et de leur faible taux d'enregistrement, les diagnostics bactériologiques du lait ne sont pas exploités en sélection.

Dans les pays scandinaves, l'enregistrement systématique des cas de mammites cliniques (MC) dans des bases de données depuis une trentaine d'années a permis d'établir des index de taureaux sur ce caractère. Les index sont définis sur la généalogie et les mesures phénotypiques de 200 filles, d'où la précision des index malgré la faible héritabilité des MC (Heringstad et al. 2000).

En France, la généralisation de l'enregistrement des MC est récente (paragraphe 1.4.2) et ce caractère n'est utilisé en sélection que depuis 2010. Mais jusqu'à présent le coefficient de détermination de cet index n'est pas très élevé. Chez les ovins laitiers et les caprins, comme la fréquence des MC est très faible (5%), aucun index sur les MC n'est disponible.

En revanche, les SCS représentent un critère de choix grâce à leur corrélation génétique très forte avec les mammites subcliniques (presque 1) et cliniques (0,7 en moyenne) et à sa facilité de mesure à grande échelle et à faible coût (paragraphe 3.2.2). Ainsi, en France et dans de nombreux autres pays, la sélection est actuellement basée sur une diminution des CCS, dans l'objectif de réduire la fréquence des mammites cliniques et subcliniques.

Les données élémentaires des CCS sont collectées dans le cadre du contrôle laitier. Chez les bovins laitiers, l'enregistrement des données de CCS a démarré au début des années 90. La base de données nationale contenait en 2010, plus de 300 millions de contrôles mensuels avec CCS, correspondant à environ 30 millions de lactations dans toutes les races

bovines laitières françaises (Rupp 2010). Chez les ovins laitiers, l'enregistrement systématique des CCS individuels a été mis en place à partir de 1999 pour les troupeaux pratiquant un contrôle laitier qualitatif officiel. Il repose sur une procédure simplifiée d'enregistrement. L'information recueillie consiste en 2 à 4 mesures individuelles de CCS par lactation, collectées après une période d'allaitement de 25 jours (Rupp et al. 2002). En 2010, la base nationale ovine de données contenait près de 2 millions d'enregistrements élémentaires de CCS issus de plus de 700 000 lactations de brebis de race Lacaune et des races ovines laitières des Pyrénées (Rupp 2010).

3.2.3.1 Evaluation génétique : calcul de l'index SCS

L'indexation consiste à estimer la valeur génétique des reproducteurs en utilisant toute l'information phénotypique des SCS valide et les relations de parenté (généalogies) disponibles dans la base de données génétique à l'aide du modèle BLUP (*Best Linear Unbiased Prediction*). La première indexation française sur les CCS a été réalisée en 1997 chez les bovins laitiers (Boichard et Rupp 1997) et en 2002 chez les ovins laitiers (Rupp et al. 2003). Elle existe chez la chèvre à titre expérimental (Rupp et al. 2011) et son officialisation est à l'étude. La sélection porte actuellement sur un critère moyen par lactation (LSCS), qui est calculé comme la moyenne arithmétique pondérée des SCS élémentaires ajustés pour le stade de lactation (Rupp et al. 2002). L'ensemble des facteurs correctifs vise à corriger la variabilité dans le nombre et la répartition (dans le temps) des SCS élémentaires utilisés pour calculer le LSCS, en particulier pour les lactations non encore terminées. Chez les bovins, il est basé sur les trois premières lactations alors que chez les ovins, il n'utilise l'information que des deux premières lactations.

Le modèle d'analyse de l'index individuel considère l'effet du troupeau, de l'année et de l'âge des animaux. Les index SCS sont exprimés en écart-type génétique, avec une inversion du signe par rapport au SCS de telle sorte que les index prennent des valeurs principalement comprises entre -3 et 3 et que les index positifs sont favorables, c'est-à-dire qu'ils correspondent à des SCS plus faibles.

La sélection a permis une amélioration de 0.5 unité de l'index SCS entre les jeunes taureaux de 2001 et ceux de 2003 (d'après Rupp, 2010b). Mais pour les brebis laitières, l'enregistrement des données CCS et l'exploitation des SCS sont trop récents pour observer une évolution.

Actuellement, le modèle d'indexation repose sur la moyenne des SCS par lactation. Or des études sont en cours pour prendre en compte toutes les valeurs mensuelles dans un modèle 'contrôles élémentaires' qui serait plus précis mais plus complexe.

3.2.3.2 Utilisation de l'index SCS dans les programmes d'amélioration génétique

La sélection doit permettre de répondre aux différentes attentes des éleveurs, à savoir améliorer la productivité laitière tout en diminuant l'impact des IMI. Ainsi les objectifs de sélection reposent sur une combinaison linéaire de l'ensemble des caractères d'intérêt pondérés par leur importance économique.

Dans les races bovines laitières, l'index de synthèse ISU est calculé à partir de l'index de sélection laitière (INEL), de l'index SCS et de trois autres index fonctionnels (morphologie, fertilité femelle et longévité fonctionnelle). Le poids de l'index SCS par

rapport à l'INEL est de 0,25 (Rupp 2010). L'introduction du nouvel index « mammites cliniques » devrait permettre d'améliorer la réponse sur la fréquence des cas cliniques.

En brebis laitière (race Lacaune), un premier index de synthèse intégrant les cellules et les caractères de morphologie a été défini en 2005 (Barillet et al. 2007). Depuis 2007, le poids de 0,5 accordé aux caractères fonctionnels (0,5xCEL et 0,5xMorphologie) est équivalent au poids accordé aux caractères de production (0,5 également) dans un index global appelé ISOL (Index de Synthèse Ovins Lait). Les objectifs de cette sélection sont de réduire d'un écart type génétique du caractère LSCS en 10 ans (Rupp 2010). Chez les ovins, l'enregistrement systématique des réformes pour cause de mammites a démarré en 2003 et devrait permettre d'évaluer l'évolution du taux de réforme liée aux infections mammaires.

3.2.4 Les apports de la génétique moléculaire

Pour améliorer la sélection sur index, une approche complémentaire commence à être utilisée. Elle repose sur l'identification de zones chromosomiques polymorphes fortement associées à un caractère et sur l'incorporation de cette information dans le choix des reproducteurs.

Dans cette partie, je vais tout d'abord définir le polymorphisme de l'ADN et donner les caractéristiques du génome ovin, puis je vais présenter les différentes régions chromosomiques associées aux mammites et enfin je donnerai quelques éléments pour l'utilisation de cette information moléculaire dans les programmes de sélection.

3.2.4.1 Le polymorphisme de l'ADN

Le génome du mouton domestique [*Ovis aries*]

Le support de l'information génétique est l'ADN (Acide Désoxyribo Nucléique). Il est composé de zones codantes (les gènes) qui représentent environ 1,5% du génome avec un ou plusieurs exons qui peuvent être traduits et des introns qui sont éliminés. Parfois un épissage alternatif entraîne la production de différents transcrits à partir d'un même gène. En amont des séquences codantes se situent une zone promotrice qui commande la transcription d'un gène ou de plusieurs gènes, on parle alors de clusters de gènes. Le reste de la molécule d'ADN se répartit entre les nombreuses zones régulatrices (cis régulation, trans régulation, micro ARN...) et de zones non codantes dont le rôle biologique n'est pas toujours connu.

Les mammifères sont des **espèces diploïdes** ie l'ensemble des molécules d'ADN est présent en deux exemplaires ; les animaux héritent d'un chromosome du père et d'un chromosome de la mère. Le nombre de chromosomes est très variable d'une espèce à l'autre. L'espèce *Ovis aries* possède 26 paires de chromosomes autosomaux et 2 de chromosomes sexuels, la vache 29 paires de chromosomes autosomaux, la souris 19 et l'homme 22.

Le génome d'*Ovis aries* est composé d'environ 3 milliards de paires de bases (3 giga bases) et de 20 à 30 000 gènes. Si la séquence bovine a été publiée pour la première fois en 2004 (Hampton 2004), puis complétée en 2009 (The Bovine Genome Sequencing and Analysis Consortium et al. 2009), la séquence ovine n'est disponible que depuis 2010. La première version se basait sur une comparaison avec la séquence bovine, puis la deuxième version se basait sur un assemblage *de novo* de la séquence ovine (The International Sheep Genomics Consortium et al. 2010) disponible sur le site internet de <http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/> Actuellement, le positionnement des gènes sur le génome ovin est dérivé de l'annotation bovine et de la

comparaison des séquences génomiques ovine et bovine. Dans cette version du génome, 20199 gènes sont annotés. Une comparaison entre les génomes ovin et bovin montre que 90% des séquences sont communes et que les séquences des exons des ovins sont plus courtes que celles des bovins (Jiang et Dalrymple 2011).

Les différentes catégories de polymorphismes

Il existe des variations de la séquence d'ADN à un locus particulier, qui se traduisent par la présence de **plusieurs allèles (polymorphisme)**. Le polymorphisme est généré par des **mutations**. Si une mutation apparaît dans les cellules germinales, elle pourra être transmise à la descendance en fonction des aléas de la méiose. Cette transmission est sous l'influence de la sélection naturelle, la sélection volontaire, la dérive génétique et les migrations de population. Ainsi, les mutations favorables aux animaux porteurs se diffusent dans la population tandis que des mutations très défavorables voire létales sont éliminées. La fréquence des mutations est extrêmement faible (10^{-6}), mais après des milliers de générations, il existe de nos jours de nombreuses séquences d'ADN polymorphes. Ces polymorphismes peuvent être de différente nature. Ils peuvent être situés sur une séquence d'ADN formée par une répétition continue de motifs composés de 1 à 10 nucléotides (**microsatellite**) ou n'être situés que sur une seule base (**polymorphismes d'un seul nucléotide, Single Nucleotide Polymorphism, SNP**). Les SNP sont présents toutes les 100 à 300 paires de bases sur le génome humain. D'autres formes de polymorphisme sont dues à une **variabilité du nombre de copies** d'un fragment d'ADN (**CNV**).

Même si les polymorphismes de l'ADN sont présents sur l'ensemble du génome, il existe des zones où leur présence est plus fréquente. Les microsatellites sont, par exemple, souvent situés dans des introns ou des exons. L'objectif de l'amélioration génétique est d'identifier les polymorphismes de l'ADN qui sont favorables à un caractère et de les sélectionner.

3.2.4.2 Les QTL associés à la résistance aux mammites

Grâce à l'évolution des techniques de génétique moléculaire des données s'accumulent sur l'existence de QTL (*Quantitative Trait Loci* QTL) associés à la résistance aux mammites. On peut dorénavant considérer que certains gènes ont des effets légèrement supérieurs sur le caractère étudié et permettent de déterminer une part non négligeable de la variance génotypique d'un phénotype. Ces QTL sont associés aux SCS ou aux mammites cliniques. Ils ont été résumés dans plusieurs ouvrages et synthèses (Rupp et Boichard 2003; Khatkar et al. 2004; Smaragdov 2006; Ogorevc et al. 2009; Rupp et Foucras 2010). Chez les bovins, 80 QTL ont été mis en évidence. Ils sont situés sur l'ensemble des chromosomes autosomes à l'exception de BTA12, 17 et 28 et ils sont plus nombreux sur BTA5, 6, 7, 9, 11, 14, 15, 18 et 23 (Rupp et Foucras 2010). En particulier, un QTL associé au SCS a été identifié sur BTA18 par une équipe allemande (Kühn et al. 2003), qui a initié un protocole de sélection de vaches sur ce QTL (Kuhn et al. 2008). Pour comprendre les mécanismes de réponse des vaches avec l'allèle favorable et de celles avec l'allèle défavorable, ils ont cultivé *in vitro* des cellules épithéliales mammaires primaires qu'ils ont ensuite infectées soit avec *E. coli*, soit avec *S. aureus* et ils ont tout d'abord mesuré l'expression de quelques cytokines et chimiokines par qPCR (Griesbeck-Zilch et al. 2009), puis de nombreux gènes par microarray (Brand et al. 2011) sur une cinétique de 24h. Les résultats de ces études seront régulièrement comparés aux résultats que nous avons obtenus.

En France entre 15 et 30 QTL associés à la résistance aux mammites ont été identifiés dans les trois principales races laitières Holstein, Normande et Montbéliarde (Fritz et al. 2008). Chez les brebis laitières, plusieurs QTL pour les SCS ont été publiés dans le cadre de protocoles basés sur l'utilisation de marqueurs microsatellites en France et en Espagne (Rupp et al. 2003; Gutiérrez-Gil et al. 2007) (Tableau 5, Rupp et al. 2010). Depuis 2010, de nouveaux projets nationaux et européens (SheepSNPQTL, 3SR : Sustainable Solutions for Small Ruminants, <http://www.3srbreeding.eu/Home.aspx>) sont en cours en France, Espagne et Italie, dans trois races ovines locales, pour rechercher de nouveaux QTL en utilisant la puce SNP 50k de cette espèce (ovine SNP Beadchip Illumina®). Ces projets s'intéressent principalement à la recherche de QTL associés aux maladies (parasitose et mammites) et à la reproduction.

Certains QTL sont spécifiques à une race, d'autres à un pathogène et d'autres encore à un phénotype particulier, ce qui explique le grand nombre de QTL détectés pour la résistance aux mammites (Rupp et al. 2010).

Tableau 5 : QTL pour la résistance aux mammites (SCS) chez les brebis laitières (revue de Rupp et Foucras 2010)

OAR	Trait	Position (cM)	Corresponding BTA region (cM)	Significance	Reference
6	SCS (L2)	80	BTA6 (65)	*	Rupp <i>et al.</i> (2003) ^a
	SCS	84	BTA6 (70)	*	Rupp (unpublished data) ^b
12	SCS (L3)	6	BTA16 (13)	**	Rupp <i>et al.</i> (2003) ^a
13	SCS (L1)	28	BTA13 (25)	*	Rupp <i>et al.</i> (2003) ^a
14	SCS	67	BTA18 (50)	*	Rupp (unpublished data) ^b
16	SCS	24	BTA20 (20)	*	Rupp (unpublished data) ^b
18	SCS (L1)	110	BTA21 (90)	*	Rupp <i>et al.</i> (2003) ^a
	SCS	72	BTA21 (60)	*	Rupp (unpublished data) ^b
20	SCS	103	BTA23 (70)	**	Gutierrez-Gill <i>et al.</i> (2007) ^c

Le but ultime de la recherche de QTL est d'identifier la mutation causale responsable du phénotype observé. Jusqu'à présent pour la résistance aux mammites, une seule mutation causale a été complètement caractérisée. Il s'agit d'un polymorphisme sur le gène FEZL (*forebrain embryonic zinc finger-like*) localisé sur le chromosome BTA22 (gène situé sur la région équivalente ovine : OAR1) dans une zone QTL associée au SCS et qui contrôle l'expression de cytokines (Sugimoto et al. 2006).

Des résultats prometteurs sont attendus des nouvelles techniques de génotypage à haut débit et de l'utilisation du déséquilibre de liaison pour établir des cartographies très fines des QTL et découvrir des mutations causales. Mais même si les mutations causales ne sont pas détectées, l'information des QTL peut être directement utilisée en sélection si les marqueurs sont suffisamment proches. Les nouvelles puces (Infinium Bovine SNP50 BeadChip d'Illumina ou son équivalent chez l'ovine) représentent un outil précieux. Chez les caprins une puce 50k est en cours de construction (INRA, International Goat Genome Consortium).

3.2.4.3 Utilisation de l'information moléculaire

Les QTL associés aux SCS dans les races laitières bovines en France ont été intégrés dans un premier programme de sélection assisté par marqueur (SAM) (Fritz et al. 2008). Les jeunes taureaux ont été évalués sur les index classiques ainsi que sur l'information apportée

par les marqueurs moléculaires et les informations sur ascendance (performances des apparentés). Ainsi la valeur génétique prédite est plus précise qu'avec la seule information sur ascendance. L'objectif de la SAM était d'utiliser l'information moléculaire disponible très précocement dans la vie du reproducteur, de façon à diminuer les coûts des schémas de sélection pour au moins une efficacité équivalente en terme de progrès génétique.

Ensuite, la création de la puce SNP bovine 50K a permis d'engager un nouveau programme de cartographie fine des QTL connus, ainsi que la détection de nouveaux QTL (Fritz et al. 2008). Ce projet a permis d'identifier des marqueurs SNP en déséquilibre de liaison avec les QTL, ce qui simplifie l'intégration de l'information moléculaire dans l'évaluation précoce des reproducteurs. Ainsi, un second programme SAM a démarré en 2008 avec l'utilisation de ces résultats et le génotypage en routine des jeunes animaux sur les 54000 marqueurs SNP couvrant le génome bovin. Les progrès génétiques gagnés par l'utilisation de l'information moléculaire n'ont pas encore été mesurés. Chez les ovins, l'information moléculaire devrait être prise en compte une fois qu'elle sera connue et fiable.

Les progrès technologiques en termes de génotypage et de cartographie de génome permettent de considérer l'ensemble des gènes impliqués dans la variabilité génétique d'un caractère d'intérêt par des marqueurs moléculaires. La sélection génomique repose sur l'estimation de l'effet des marqueurs sur des caractères d'intérêt, qui est établie dans une population de référence. Puis ces modèles de prédiction sont appliqués à la jeune génération de reproducteurs afin d'indexer cette population avant de connaître les performances de ces reproducteurs. Ainsi la sélection génomique permet d'améliorer les schémas de sélection surtout pour les caractères qui sont peu héréditaires comme les SCS ou les MC, car elle permet d'accélérer le processus de sélection des reproducteurs (index précis et précoces) avant les résultats de performance des jeunes animaux. Toutefois, il est important de ré-estimer régulièrement les effets des marqueurs sur les performances des animaux les plus jeunes, sans quoi la précision de l'estimation « sélection génomique » diminue rapidement au cours des générations. Chez les ovins et les caprins des programmes sont en cours pour évaluer la faisabilité et les modalités de la sélection génomique dans ces espèces à partir de populations de référence de taille suffisante.

En France, la sélection basée sur les index SCS est opérationnelle chez les principales races laitières bovines et ovines et à l'étude chez les races caprines. Des diminutions des SCS et de la fréquence des IMI sont attendues. L'utilisation de l'information moléculaire des QTL en SAM peut améliorer l'efficacité de la sélection. Toutefois le nombre important de QTL détectés confirme la nature complexe et polygénique de la résistance aux mammites et rend plus difficile la mise en place de la SAM. Actuellement la sélection génomique qui prend en compte l'intégralité de l'information aux marqueurs est prometteuse. Mais pour sa bonne efficacité, il est important de réévaluer régulièrement les liens entre les SNP et les phénotypes.

De plus, la sélection sur les index ou sur des marqueurs moléculaires fonctionne comme une « boîte noire » car les mécanismes sélectionnés ne sont pas connus. Une meilleure connaissance de la base génétique du caractère permettrait de mieux comprendre et évaluer les mécanismes en jeu lors de la sélection sur les SCS ou les MC, ces caractères étant indirects et sans doute imparfaits. Elle permettrait aussi de mieux évaluer les conséquences de cette sélection à long terme et d'appréhender une dynamique hôtes-agents pathogènes. La recherche de QTL associé à la résistance aux mammites pourrait permettre d'identifier des mutations ou des gènes candidats, ce qui permettrait une sélection directe plus efficace et plus durable que la SAM ou la sélection génomique. La recherche de QTL pourrait éventuellement être appliquée à d'autres phénotypes que les SCS. Mon travail de thèse s'inscrit dans cette

démarche de mieux connaître les mécanismes de réponse des ruminants laitiers aux IMI, en utilisant des nouveaux outils de la génétique moléculaire fonctionnelle.

4 Objectifs de la thèse

Comme nous l'avons vu précédemment, les IMI constituent une pathologie majeure en élevage laitier. Par ailleurs la sélection opérée à partir des années 1970 jusqu'à nos jours pour une meilleure productivité laitière s'est accompagnée d'une augmentation de la sensibilité aux mammites chez les bovins avec de fortes conséquences économiques (diminution du volume de lait, pénalité sur le prix du lait vendu, traitement des animaux infectés, prévention...). Cette dégradation de la santé de la mamelle a conduit à considérer les IMI dans les objectifs de sélection des races laitières. Les CCS sont les témoins à la fois de la présence d'une infection et de la présence d'une réponse inflammatoire plus ou moins intense. Ainsi, une mesure des CCS sur un rythme mensuel est un très bon indicateur de la présence d'une IMI.

En France, les objectifs reposent principalement sur une diminution des CCS et récemment, chez les bovins uniquement, sur la prise en compte des MC. L'indexation sur les SCS a commencé en 1997 pour les bovins laitiers et en 2002 pour les ovins laitiers. Elle commence à donner de bons résultats car l'index SCS des jeunes reproducteurs s'améliore légèrement (+0,5 entre 2001 et 2003 chez les bovins).

Chez les bovins, la détection de QTL associés aux MC ou aux SCS a permis de compléter la sélection sur l'index SCS par une SAM. Les avancées technologiques du génotypage et les baisses de coût ont permis d'envisager une sélection génomique, c'est-à-dire une sélection qui prend en compte les déséquilibres de liaison de dizaines de milliers de marqueurs avec les QTL. La sélection génomique a le fort avantage de permettre une sélection précise et précoce sur des caractères qui ont des héritabilités faibles, voire très faibles comme les SCS ou les MC. Des études sont en cours pour prendre en compte ces mêmes critères en élevage ovin laitier, principalement dans la race Lacaune.

Cependant, la sélection génétique quantitative classique ou avec des marqueurs moléculaires, pour une diminution des SCS, repose sur une « boîte noire ». En effet à l'exception d'un QTL associé au SCS dont le rôle a été détaillé (Sugimoto et al. 2006), les mécanismes génétiques de résistance aux mammites ne sont pas connus. Ainsi, la sélection ne considère sûrement pas la diversité des agents pathogènes responsables de mammites ni toute la complexité du caractère de résistance.

L'étude bibliographique au chapitre I a montré que de nombreuses informations sur la réponse des différents types cellulaires (CSL, CEM et plus récemment macrophages) à des infections principalement à *S. aureus* et *E. coli* sont disponibles. En revanche ces analyses comparent généralement des échantillons sains et à des échantillons infectés ou élaborent des cinétiques de réponse aux infections, mais ne prennent pas en compte la variabilité de la réponse de l'hôte.

C'est pourquoi des scientifiques de l'INRA de Toulouse ont mis au point un protocole de sélection divergente basée sur les SCS depuis 2003. Même si les bovins occupent une place dominante avec trois races principales en France (Prim'Holstein, Normande et Montbéliarde), un modèle ovin a été choisi. Les ovins ont l'avantage d'être proches phylogénétiquement des bovins avec des physiopathologies assez comparables, même si la fréquence des MC est nettement plus faible chez les ovins. Ainsi les résultats obtenus chez les ovins peuvent être extrapolés, au moins en partie, aux bovins. De plus, la durée des intervalles entre générations chez les ovins est plus courte et à nombre d'animaux identiques, leur

élevage est moins coûteux (taille plus petite des bâtiments, quantité moindre d'aliments consommés,...) que chez les bovins, ce qui a favorisé le choix des ovins comme espèce modèle.

Ce protocole de sélection divergente a permis de générer deux lignées de brebis Lacaune, une avec des SCS élevés et une autre avec des SCS plus faibles. Les objectifs précis de ce protocole étaient multiples :

- 1- apprécier les conséquences d'une sélection sur les SCS sur la capacité de la mamelle à répondre aux infections ;
- 2- déterminer les mécanismes régulant la réponse immunitaire et inflammatoire ;
- 3- connaître les gènes ou les régions chromosomiques impliqué(e)s dans la sélection sur les SCS ;
- 4- savoir si la sélection intense sur les SCS peut provoquer un risque accru de sensibilité à l'infection.

Mon travail de thèse s'inscrit principalement dans les objectifs 1, 2 et 3 de ce protocole. Ainsi, les objectifs de mon travail de thèse étaient :

- de caractériser les gènes et mécanismes mis en place lors d'une infection staphylococcique dans les deux lignées de brebis par des analyses d'expression des gènes à l'aide de la technologie microarray ;
- de comparer les résultats obtenus au sein de chaque lignée pour identifier les conséquences de la sélection sur les SCS ;
- d'identifier des polymorphismes de l'ADN responsables des (ou de certains) mécanismes de résistance.

Notre modèle reposant sur des ovins, il n'était pas pertinent d'étudier la réponse à *E. coli* qui est un agent pathogène très peu fréquent chez la brebis (paragraphe 1.2) et nous nous sommes concentrés sur la réponse à *S. aureus*.

Pour caractériser les gènes et mécanismes de réponse des brebis à des infections, nous avons élaboré trois études différentes. Tout d'abord, nous avons étudié *in vivo* la réponse des **cellules somatiques du lait (CSL)**, principalement composées de neutrophiles, après inoculation de *S. epidermidis* et de *S. aureus* (Chapitre III, Partie 1). Nous avons choisi ces deux espèces bactériennes, car elles sont principalement responsables de mammites subcliniques et cliniques chez les ovins, et qu'elles permettent donc d'étudier la réponse des hôtes en interaction avec un spectre plus divers d'agents pathogènes. Ensuite, nous avons étudié les lignées divergentes *in vitro* à travers deux autres types cellulaires, supposées jouer un rôle dominant, i) nous avons choisi les **cellules dendritiques (CD)** pour leur rôle central dans l'immunité, avec leurs capacités de détecter les antigènes et de transmettre l'alerte aux autres cellules du système immunitaire (Chapitre III, Partie 2), ii) nous nous sommes intéressés aux **cellules épithéliales mammaires (CEM)** car elles forment une barrière physique entre la lumière et les tissus de la glande mammaire et peuvent sécréter des cytokines chimio-attractantes pour diverses cellules (Chapitre III, Partie 3). Ensuite, une méta-analyse de ces trois jeux de données a permis de confirmer les hypothèses qui avaient été formulées dans chaque analyse individuelle, et d'en générer de nouvelles. Elle a aussi permis d'établir une liste de mécanismes et de gènes candidats fonctionnels pour déterminer l'origine génétique de la différence entre lignées. Cette synthèse est présentée dans la Partie 4 du Chapitre III.

Pour compléter l'étude génétique des lignées, nous avons réalisé des études du polymorphisme de l'ADN sur ces animaux (Chapitre III, Partie 5). Dans un premier temps

nous avons étudié séparément les deux lignées de brebis pour essayer de valider ou découvrir de nouvelles zones chromosomiques qui jouent un rôle dans la sélection sur les SCS. Nous nous sommes aussi concentrés sur un deuxième phénotype, la présence d'abcès mammaires pour lequel nous possédions un phénotypage fin des animaux. Ensuite les résultats des analyses fonctionnelles et positionnelles ont été confrontés pour mettre en évidence les principales caractéristiques des brebis de la lignée résistante et des brebis de la lignée sensible aux IMI.

CHAPITRE II : MATERIELS ET METHODES

Dans ce deuxième chapitre, je vais commencer par présenter les particularités biologiques des brebis sur lesquelles j'ai travaillé en expliquant le protocole de génération de deux lignées divergentes de brebis sélectionnées sur le SCS. Je vais montrer les principaux résultats en infections naturelles et expérimentales déjà connus avant le début de ma thèse. Ensuite, je vais présenter les principaux outils qui permettent de mettre en évidence des différences d'origine génétique entre les deux lignées de brebis. Je vais principalement développer les études du transcriptome avec la puce à ADN et dans une moindre mesure les études du polymorphisme de l'ADN. Ce chapitre n'a pas pour vocation de donner les éléments précis de méthodologie qui permettraient de reproduire les expériences (ces données étant présentées dans les parties « Matériel et Méthodes » de chaque partie du Chapitre 3), mais de faire un état de l'art sur les différentes approches.

1 Un modèle d'étude original : les lignées divergentes

Avec l'utilisation de plus en plus répandue du SCS en sélection chez les ruminants laitiers des questions ont émergé et ont conduit des chercheurs de l'INRA de Toulouse à mettre au point une sélection divergente de ruminants basée sur les SCS afin de connaître les conséquences à long terme d'une telle sélection. Un modèle animal ovin a été choisi car l'élevage d'ovins est moins coûteux que celui de bovins et que la réponse aux IMI est proche chez les trois principaux ruminants laitiers (bovin, ovin, caprin), même si la fréquence des MC est moindre chez les ovins.

1.1 Protocole de sélection divergente

1.1.1 Production des brebis des lignées divergentes

Des brebis de race Lacaune ont été procréées au domaine expérimental INRA de La Fage (UE321 Roquefort). Le protocole a consisté en une sélection divergente sur une génération basée sur les index SCS des parents (Rupp et al. 2009). La première cohorte a été créée en 2004.

Pour la production des lignées divergentes, deux groupes de six à sept béliers testés sur descendance, provenant des deux unités de sélection Confédération Générale de Roquefort et Ovitest, et présentant des index SCS extrêmes sont choisis chaque année. Ce choix se fait annuellement depuis 2003 suite à une indexation (Rupp et al. 2009; Allain et al. 2010). Les béliers pères des brebis CCS+ ont des index SCS négatifs (-1,96 en moyenne), tandis que les béliers pères des brebis CCS- ont des index SCS positifs (1,43 en moyenne). Les béliers sont aussi choisis pour avoir des index précis (coefficients de détermination 0,85 en moyenne) et des valeurs de l'index du lait positives et équivalentes.

Ces mâles sont accouplés par insémination artificielle(IA) avec les brebis de l'UE de La Fage pour produire deux groupes de filles avec des CCS respectivement élevées (CCS+) et faibles (CCS-). Une intensité de sélection moindre est appliquée aux mères, qui sont réparties dans deux groupes de même effectif d'après leur « cellules ». Depuis 2007, des brebis de 2^{ème} génération sont également produites.

Les protocoles successifs de lignées divergentes de première génération entre 2005 et 2009 ont abouti à la mise en lactation de 191 brebis CCS+ et de 210 brebis CCS- (Allain et al. 2010).

1.1.2 Des différences de concentrations de cellules somatiques

Pour caractériser les brebis des deux lignées divergentes, plusieurs mesures ont été effectuées. Elles ont consisté à :

- mesurer les caractères laitiers et les CCS de toutes les brebis pendant l'allaitement puis toutes les 3 semaines dans le cadre du contrôle laitier, soit 9 fois entre les mois de janvier et de juin, et 7 fois pour les primipares, leur lactation étant plus courte avec une IA un mois après les multipares ;
- enregistrer les cas de MC ;
- réaliser des examens cliniques des mamelles par inspection et palpation pour détecter la présence d'abcès à quatre reprises par lactation ;
- analyser les bactériologies du lait : une fois au moment de la mise bas et trois fois pendant la lactation (hors période d'allaitement).

Les lignées créées ont permis d'obtenir une divergence de 1,75 point de score, ce qui correspond à 3 écart-types génétiques du SCS et à plus du double de la moyenne géométrique des CCS entre les lignées CCS+ et CCS- pour les lactations de rang 1 à 3 dès la première génération. Ce résultat s'illustre par des moyennes brutes égales à 279 000 cellules/ml dans la lignée CCS- et à 1 209 000 dans la lignée CCS+ (Allain et al. 2010) (Tableau 6).

Tableau 6 : Caractéristiques des brebis CCS+ et CCS- (Lactations 1 à 3, Traite du matin, 2005 à 2009) (Allain et al. 2010)

	Lignée CCS+	Lignée CCS-
Effectif	191	210
CCS (*1000 c/ml)	1209 (\pm 2955)	279 (\pm 1289)
SCS	4,4 (\pm 2,4)	2,6 (\pm 1,7)

1.2 Résistance aux infections intra-mammaires dans les lignées divergentes

1.2.1 Suivi d'infections naturelles

Comme nous l'avons vu précédemment, la prévalence des mammites subcliniques est généralement élevée (Tableau 6) ; il en est de même dans le troupeau de l'unité expérimentale. Les brebis CCS+ et CCS- sont donc exposées en continu à des agents responsables de mammites.

Le suivi des brebis CCS+ et CCS- pendant 5 années a montré qu'une sélection sur les SCS est favorable à la résistance aux IMI naturelles (Tableau 7) (Rupp et al. 2010; Allain et al. 2010). En effet, même si peu de cas de MC sont survenus, la plupart est apparue chez les brebis CCS+ (n=25 cas sur 31, fréquence globale de 7%). De plus, l'incidence des mammites chroniques associées à des abcès détectés par l'examen mammaire était significativement supérieure chez les brebis CCS+ que chez les brebis CCS- (36% contre 9% de brebis affectées).

Les analyses bactériologiques du lait montrent que la prévalence des IMI est plus élevée chez les brebis CCS+ que chez les brebis CCS- (OR=3,3[2,5-4,4]), en particulier au moment de la mise bas (OR=7,0[4,2-11,7]) (Rupp et al. 2010; Allain et al. 2010). De plus, l'identification précise des agents pathogènes a été effectuée par bactériologie sur les échantillons de la première cohorte (Rupp et al. 2009). Ainsi, la durée des infections est plus

longue chez les brebis CCS+ que chez les brebis CCS-. Les staphylocoques sont les agents pathogènes majeurs, particulièrement *S. auricularis* avec 20,8% des échantillons de lait positifs chez les brebis CCS+ et 5% chez les brebis CCS-. Les prévalences de *S. aureus* et *S. epidermidis* sont respectivement de 1,2% et 0,3% chez les brebis CCS+ et de 0,6 et 0,4% chez les brebis CCS- (Rupp et al. 2009).

Tableau 7 : Concentrations de cellules somatiques et fréquences des mammites chez les brebis CCS+ et CCS- (Allain et al. 2010)

	Lignée CCS+		Lignée CCS-	
	N	Moyenne	N	Moyenne
Mammites cliniques (762 L1-L3)	355	25 cas	407	6 cas
Abcès (au moins 2 palpations)				
% de L1-L3 affectées	323	36 %	382	9 %
Bactériologies du lait (2005-2007)				
% de positif sur tous les échantillons ¹	1133	46 %	1294	23 %
% de positif à la mise bas	218	54 %	248	22 %

¹ échantillons prélevés à la mise bas, pendant l'allaitement et lors de 3 contrôles laitiers espacés de 6 semaines

La sélection divergente sur les SCS est équivalente à environ 30 ans de sélection suivant l'objectif de sélection des ovins laitiers proposé pour la race Lacaune en 2006. Tous les résultats sont publiés dans l'article décrivant les lignées (Rupp et al. 2009). Les observations en ferme ont confirmé le fait que les brebis sélectionnées pour des SCS élevés sont plus sensibles aux infections mammaires que les brebis sélectionnées pour des SCS faibles.

1.2.2 Suivi d'infections expérimentales

Pour caractériser les différences de réponse aux IMI des brebis CCS+ et CCS- des infections expérimentales ont été réalisées à l'école nationale vétérinaire de Toulouse (ENVT). Ces expériences, basées sur un faible nombre d'animaux ont permis d'évaluer la réponse des brebis CCS+ et CCS- dans des conditions contrôlées d'infection (agent pathogène, quantité inoculée, début de l'infection).

Sept brebis CCS- et sept brebis CCS+ de la première cohorte ont été choisies car elles n'avaient aucun signe préexistant d'IMI (SCS faible, bactériologie du lait négative). Ces brebis n'avaient aucun lien parental direct entre elles. Elles ont été inoculées expérimentalement par *S. epidermidis* en 2005, puis par *S. aureus* en 2006. Ces deux staphylocoques sont les principaux agents responsables de mammites respectivement subcliniques et cliniques, voire chroniques chez les ovins. Les brebis ont été suivies pendant les 10 jours suivant l'inoculation. Les résultats acquis avant ce travail de thèse ont été synthétisés par Rupp al. (2010) et en partie présentés dans Bonnefont et al. (2011).

En résumé, après chacune des inoculations, les SCS et les profils d'excrétion bactérienne (Figure 3) ne sont pas significativement différents sur l'ensemble de la période observée. En revanche, après inoculation de *S. aureus*, les excréments bactériennes dans le lait étaient plus élevées sur la période 24-48h chez les brebis CCS+ que chez les brebis CCS-

($\mu_{CCS+}=13,7$ (1,7 écart-type) et $\mu_{CCS-}=12,2$ (2,1), $p=0,048$) et après inoculation de *S. epidermidis*, les excrétions bactériennes avaient tendance à être plus élevées à 48h chez les brebis CCS+ que chez les brebis CCS- ($\mu_{CCS+}=3,9$ (4,6) et $\mu_{CCS-}=0,86$ (0 ;21), $p=0,27$) (Bonnefont et al. 2011).

D'autre part, des palpations mammaires ont été effectuées au moment de la traite du matin et du soir. Les principaux résultats ont montré que des abcès mammaires étaient détectés chez quatre brebis CCS+ et une seule brebis CCS- et que le score moyen de l'examen clinique (index synthétique du nombre et de la taille des abcès mammaires) était plus élevé chez les brebis CCS+ que chez les brebis CCS- après inoculation de *S. epidermidis* ou de *S. aureus* (Figure 3).

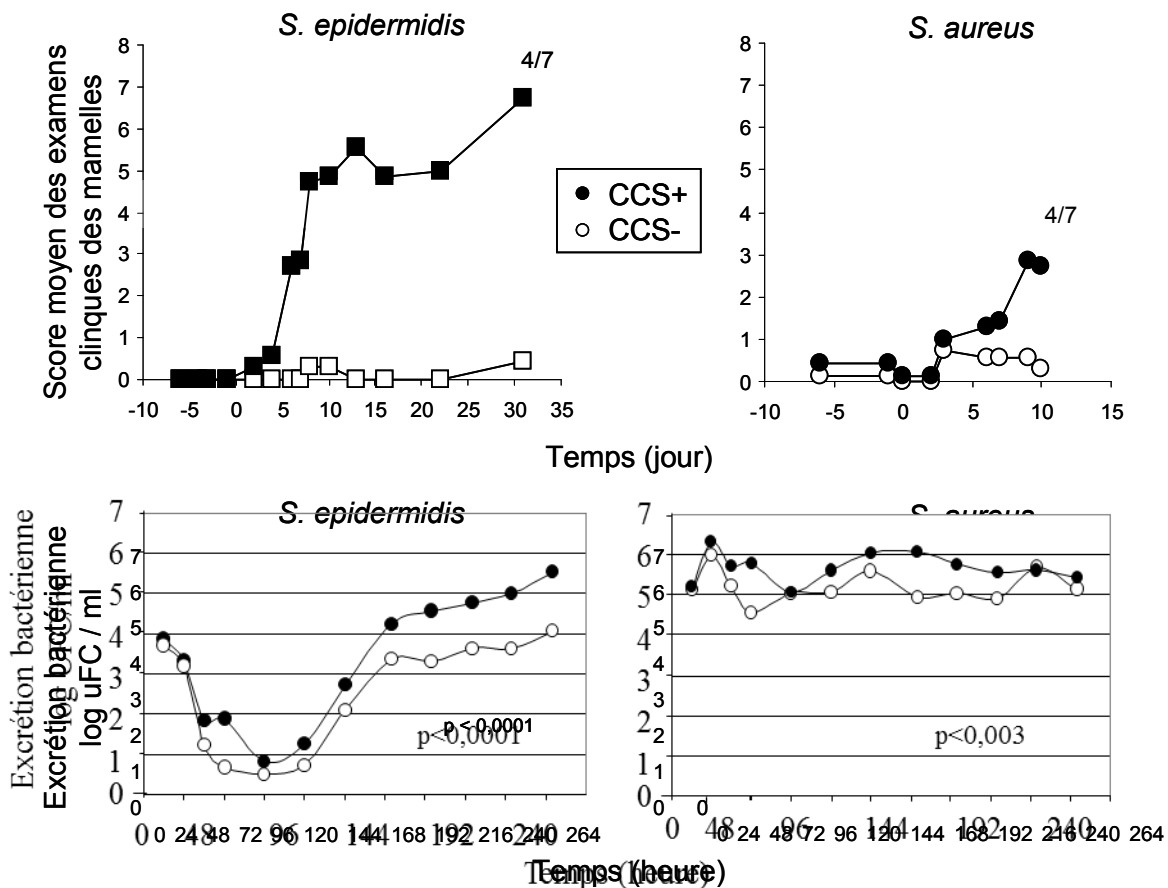


Figure 3 : Score moyen des examens cliniques des mamelles et excrétions bactériennes (log uFC / ml) après inoculation de 10^3 ufc *S. epidermidis* et de *S. aureus*. Les p-values correspondent à l'effet lignée.

Le suivi clinique des brebis après infection expérimentale par deux staphylocoques confirme les observations faites dans l'UE en situation d'infection naturelle. En effet, l'ensemble des résultats montre une sensibilité aux IMI supérieure chez les brebis de la lignée avec des CCS élevées que chez celles de la lignée avec des CCS faibles.

En conclusion, il semble que les brebis CCS- possèdent certaines aptitudes qui leur permettent de limiter la multiplication bactérienne en cas d'IMI. Nous pouvons également proposer l'hypothèse que dans un troupeau avec des brebis CCS+, la contamination des brebis entre elles est plus importante que dans un troupeau de brebis CCS- car l'excrétion bactérienne est supérieure.

2 Analyses des différences d'origine génétique

A la suite de la description des ressources animales, je vais présenter les outils qui permettent d'analyser les différences d'origine génétique. Ces différences peuvent s'appréhender à trois échelles. La première échelle est le support de **l'information génétique** : l'**ADN** (acide désoxyribo nucléique). Il a la propriété d'être **identique** dans toutes les cellules somatiques d'un organisme vivant. La deuxième exploration repose sur l'étude du **transcriptome**, c'est-à-dire de l'ensemble des gènes transcrits en **ARNm** (Acide RiboNucléique messenger). Le transcriptome d'une cellule dépend de nombreux facteurs : de sa spécialisation (musculaire, immunitaire, neurone, ...), de son état physiologique, de la présence d'hormones, de nutriments, de cytokines, de pathogènes ... Une troisième échelle serait l'étude du **protéome**, c'est-à-dire de l'ensemble des **protéines** d'une cellule. Comme précédemment pour le transcriptome, le protéome dépend fortement du type de cellule étudié et des conditions environnantes. Ces trois niveaux de l'information génétique (ADN, ARNm et protéines) interagissent dans une cellule et apportent des informations complémentaires. Au cours de ma thèse, j'ai principalement travaillé sur le transcriptome et j'ai aussi abordé l'étude de l'ADN génomique. Je ne décrirai les méthodes que de ces deux types de données.

2.1 L'analyse des gènes exprimés

On considère qu'il y a environ 20 à 30 000 gènes différents dans le génome ovin. Pour mesurer l'expression des gènes, il existe plusieurs approches. Une des principales approches est la **RT-qPCR** (transcription réverse et Réaction de Polymérisation en Chaîne quantitative). Cette approche permet de mesurer ponctuellement l'expression d'un gène. Or, dans une cellule, entre 3000 et 5000 gènes peuvent être transcrits en même temps. Ainsi pour étudier le transcriptome, il est intéressant d'utiliser un outil d'analyse de l'expression des gènes à haut débit : le **microarray** ou puce à ADN. Cette technique permet d'analyser simultanément des milliers de gènes. Une nouvelle approche basée sur la technologie de séquençage s'avère prometteuse : elle consiste à séquencer l'ensemble des molécules d'ARNm d'une cellule (**RNAseq**). Dans cette section, je présenterai brièvement la PCR et ses applications, je développerai plus en détail les analyses avec les microarrays.

2.1.1 Quelques méthodologies d'analyse

2.1.1.1 Quantification des ARN par réaction de polymérisation en chaîne

La RT-qPCR consiste à amplifier une séquence d'ADN à partir d'une amorce qui est spécifique au transcrite à mesurer. Elle repose sur des cycles de température précis qui déterminent l'enchaînement des étapes de **polymérisation** de la molécule d'ADN cible et de **dénaturation** de l'ADN double brin obtenu en ADN mono-brin. L'étape de fixation de l'amorce à l'ADN est déterminante pour l'efficacité de la réplication du transcrite. Les cycles de température du milieu de réaction sont bien contrôlés dans le thermocycleur, ce qui garantit l'efficacité de la RT-qPCR. Des marquages avec des fluorochromes permettent de quantifier les molécules d'ADN au cours du temps et donc de suivre l'amplification des transcrits. L'intensité lumineuse mesurée est proportionnelle à la quantité d'ADN synthétisée.

Le milieu de réaction contient :

- la séquence d'ADN à amplifier ;
- une ADN polymérase (enzyme capable de répliquer l'ADN, il s'agit souvent d'une TAQ polymérase : *Thermophilus aquaticus*) ;
- des amorces d'ADN spécifiques à la séquence d'ADN à amplifier ;
- des bases azotées, qui seront assemblées lors de la synthèse de nouvelles molécules d'ADN ;
- un intercalant (souvent du SYBR green) qui se fixe à l'ADN synthétisé et permet de mesurer la quantité d'ADN synthétisé.

L'amplification de l'ADN n'est pas linéaire sur l'ensemble des cycles, mais elle est souvent exponentielle, ainsi on représente souvent le logarithme de la quantité d'ADN en fonction des cycles (**Figure 4**). On utilise la valeur de la phase exponentielle (*ie*, du milieu de la phase linéaire sur le graphique en logarithme) pour comparer les niveaux d'expression entre différents échantillons. Mais cette valeur en soi n'est pas très informative. Il faut normaliser ces valeurs entre les différents échantillons en référence à des gènes de ménage (*house keeping genes*, *HKG*) dont l'expression n'est pas affectée par l'état physiologique de la cellule/du tissu et par les conditions environnementales. Actuellement, on considère qu'il faut utiliser l'expression de plusieurs gènes de ménage (souvent 4 ou 5) et exprimer la quantité d'ADN d'un gène d'intérêt en expression relative par rapport aux gènes de normalisation (Vandesompele et al. 2002).

Pour vérifier que la RT-qPCR est spécifique, la courbe de dissociation peut être tracée (**Figure 5**). Elle montre la température à laquelle les ADN double-brin se séparent en ADN monobrin. Si un seul pic est visible sur les courbes, la réaction est spécifique.

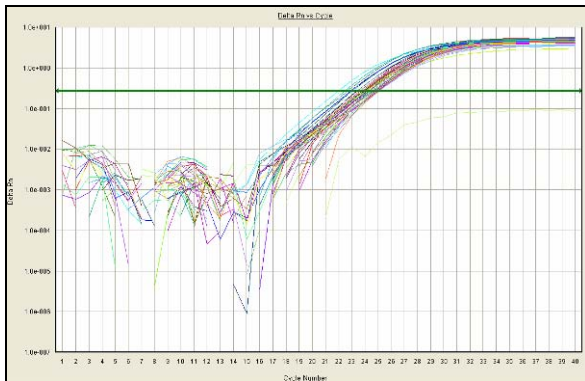


Figure 4 : Courbes de suivi de la quantité d'ADN au cours de la RT-qPCR. Exemple du gène de ménage RP19 pour les échantillons des cellules épithéliales mammaires.

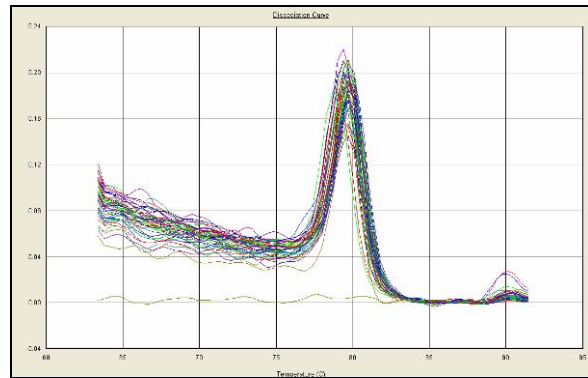


Figure 5 : Courbes de dissociation des ADN double-brin en ADN monobrin pour le gène de ménage RP19 avec les échantillons des cellules épithéliales mammaires.

Pour appliquer cette réaction à des échantillons d'ARNm, il faut effectuer une première étape appelée transcription inverse (RT), elle permet de synthétiser un premier brin d'ADN à partir d'une molécule d'ARNm. La quantité d'ADN peut alors être mesurée par qPCR.

Actuellement, il existe une nouvelle technologie Biomark® commercialisée par la société Fluidigm qui permet de mesurer de nombreux échantillons simultanément (plaque de 48*48 ou 96*96) et de réduire fortement les volumes de réaction (6.7 nl). Cette technologie permet de gagner du temps de réaction et d'économiser de grands volumes de réactifs.

2.1.1.2 Quantification des ARN par puce à ADN

2.1.1.2.1 Nature et définitions d'une puce à ADN

Spot : zone de la puce d'ADN sur laquelle sont fixés des millions de sondes identiques.

Sonde (probe) : séquence d'ADN connue fixée sur l'emplacement d'un spot de la puce à ADN.

Les microarrays sont des outils qui permettent d'étudier l'expression de plusieurs centaines, voire milliers de gènes simultanément.

Nature des microarrays

Les microarrays peuvent se présenter sous forme de lame de verre ou de membrane de nylon. Les microarrays sur lame de verre sont les plus répandus, car leur coût est relativement faible, ils sont stables à de fortes températures, peuvent être lavés et ré-hybridés et ont un bruit de fond en fluorescence très faible (Cheung et al. 1999).

Ils sont spécifiques à une espèce particulière, car leurs sondes proviennent de cette espèce. L'avantage des puces à ADN à haute densité vient de leur caractère « **pangénomique** », ce qui signifie que la majeure partie des gènes d'un organisme est représentée sur cette puce. Par conséquent, lorsqu'on utilise cette puce, on fait peu d'hypothèse *a priori* sur les gènes exprimés dans une cellule (Sobek et al. 2006).

Plusieurs techniques de dépôt des sondes

Il existe deux grands types de conception des microarrays. Initialement, les sondes étaient synthétisées puis imprimées sur chaque spot (*printed microarrays*). Actuellement, la plupart des microarrays contiennent des oligonucléotides qui sont synthétisés *in situ* sur les spots (*in situ-synthesized oligonucleotide microarrays*) (Miller et Tang 2009). Cette seconde technique a permis d'améliorer la performance des puces à ADN et de densifier énormément le nombre de spots sur la puce jusqu'à atteindre des centaines de milliers de spots sur une même lame (technologie Affymetrix). Les sondes d'ADN sont alors monobrins et n'ont pas besoin d'une étape de dénaturation. De plus, grâce à la technique de synthèse *in situ*, toutes les sondes d'une puce ont la même longueur et la même température de fusion, ce qui permet une hybridation plus homogène à une température donnée (Barrett et Kawasaki 2003).

Diverses longueurs de sondes

Avant la disponibilité des séquences d'un génome, les gènes exprimés étaient identifiés par des analyses de séquence d'ADN complémentaire (ADNc). Ces ADNc étaient insérés dans des plasmides puis amplifiés par PCR. Puis ils étaient fixés sur la lame soit de façon électrostatique (chaleur ou UV), soit par des liaisons covalentes en 5' (Barrett et Kawasaki 2003). L'utilisation d'ADNc permet une très bonne spécificité d'hybridation même sans connaître la séquence nucléotidique. Mais cette technique était coûteuse pour l'obtention de milliers de gènes et avait l'inconvénient d'avoir un bruit de fond élevé.

Avec la technique de synthèse *in situ*, la taille des séquences d'oligonucléotides a diminué. On différencie les séquences courtes qui contiennent une vingtaine de nucléotides (25 chez Affymetrix), des séquences plus longues de 50 à 80 bases (Barrett et Kawasaki

2003). Avec les séquences courtes, la spécificité d'hybridation est un peu plus faible car plusieurs cibles peuvent s'hybrider sur une même sonde ; aussi plusieurs sondes différentes sont nécessaires pour évaluer l'expression d'un même transcrite (sur les microarrays Affymetrix, chaque transcrite est représenté par 11 à 20 sondes) ce qui permet d'éliminer les problèmes de spécificité (Barrett et Kawasaki 2003). Lorsque les séquences sont un peu plus longues, la spécificité d'hybridation est meilleure. Il semble que des oligonucléotides de 60 mères soient un bon compromis entre spécificité et sensibilité (Hughes et al. 2001) et il a été montré qu'avec des oligonucléotides de 50 mères la spécificité d'hybridation est aussi bonne qu'avec des ADNc (Kane et al. 2000).

La puce ovine Agilent : ses caractéristiques

Au cours de ma thèse, j'ai utilisé la puce Agilent ovine 15K (019921). C'est une lame dédiée à l'espèce ovine. Elle contient 8 microarrays identiques avec 15 208 spots. Parmi ces spots, il y a 200 spots de contrôle et environ 12 000 gènes, soit environ 50 à 70% de la totalité des gènes ovins, dont 8 000 seulement sont actuellement annotés (Casel et al. 2009).

La technologie *inkjet printing process* utilisée pour synthétiser *in situ* les sondes confère une bonne reproductibilité à ce microarray. Les oligonucléotides comportent 60 nucléotides ; cette taille permet d'obtenir une bonne spécificité d'hybridation. Le nombre élevé de sondes permet d'évaluer l'expression d'environ un tiers des gènes de l'espèce ovine. De plus, la technologie sur lame de verre permet l'hybridation de 2 échantillons en parallèle sur le même microarray, ainsi à chaque hybridation d'une lame de 8 microarrays, 16 échantillons peuvent être analysés simultanément. La bonne qualité du dépôt des sondes donne la possibilité à l'utilisateur de dénaturer la première hybridation et d'en faire une seconde avec un autre échantillon. Toutes ces caractéristiques de la puce Agilent ovine ont contribué à notre choix de travailler sur ce microarray.

La facilité d'utilisation des microarrays et la quantité d'information obtenue par leur utilisation à un coût raisonnable a entraîné, dans les années 2000, une explosion du nombre de publications utilisant cette technique chaque année (Figure 6).

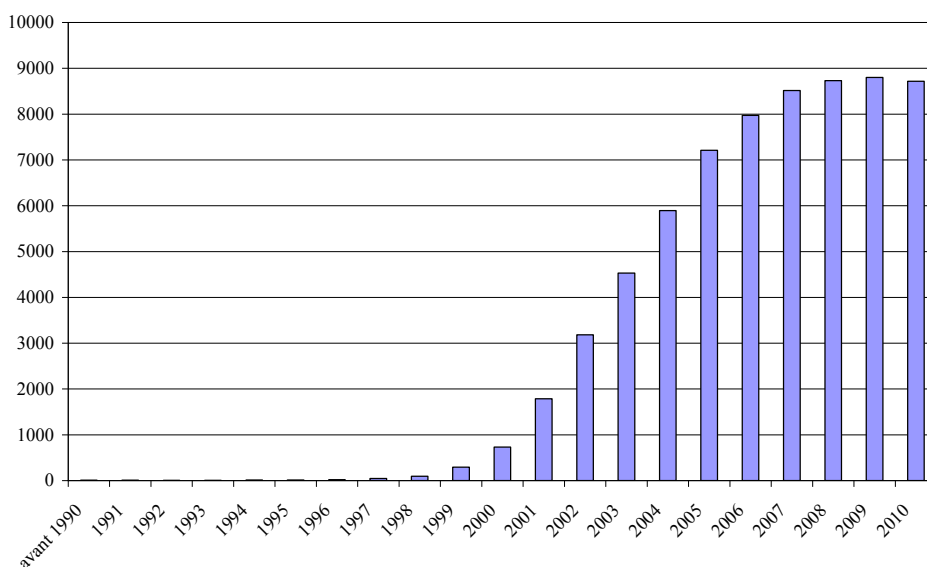


Figure 6 : Histogramme du nombre de publications utilisant la technologie microarray par an depuis 1990 (d'après the web of knowledge).

2.1.1.2.2 Hybridation des échantillons d'ARNm

Cible (target) : séquence d'ADN copie d'un échantillon dont on veut mesurer l'expression des gènes.

Hybridation (hybridisation) : action qui conduit à un attachement spécifique entre une cible d'ADN provenant d'un échantillon et une sonde d'ADN fixée sur une puce à ADN dans des conditions expérimentales précises. Pour pouvoir s'hybrider ensemble, les séquences de la cible et de la sonde doivent être complémentaires.

Préparation de la cible

Une fois que les cellules d'intérêt sont isolées, leurs ARNm sont extraits. Ces ARNm sont amplifiés en ARN copie qui constituent la **cible**. Il est important de vérifier la qualité des ARN avant l'hybridation (*RNA integrity number* : RIN) (Schroeder et al. 2006) afin de garantir la qualité des résultats. La cible est marquée soit par de la radioactivité (principalement ³³P), soit par un fluorochrome (technique la plus utilisée ; souvent Cyanine 3 –Cy3– et/ou Cyanine 5 –Cy5–). La cible est préparée selon un protocole expérimental standardisé, puis placée en incubation sur le microarray pour l'hybridation.

Habituellement, le microarray est spécifique de la cible. Toutefois, lorsqu'il n'existe pas de microarray spécifique à une espèce, il est possible de faire des hybridations hétérologues (Bar-Or et al. 2007), c'est-à-dire de choisir un microarray qui contient des sondes d'ADN d'une espèce différente. Il est alors important de choisir une espèce proche phylogénétiquement afin d'optimiser l'hybridation entre la sonde et la cible. Mais les résultats d'hybridation hétérologue sont mitigés (Bar-Or et al. 2007).

Hybridation spécifique de la cible sur la sonde

Les ADNc dont la séquence est complémentaire à celle d'une sonde s'hybride sur celle-ci. Puis un lavage du microarray permet d'éliminer les cibles non fixées. Une hybridation réussie entre une cible marquée et une sonde immobilisée engendre une augmentation de l'intensité de la fluorescence ou de la radioactivité sur un spot par rapport à un niveau environnant (bruit de fond) (Miller et Tang 2009).

Obtention du fichier de données

La lecture du microarray s'effectue avec un scanner (GenePix 4000B, Axon Instrument pour de la fluorescence ou MicroImagers, Biospace Mesures pour de la radioactivité). Elle permet de mesurer l'intensité du marquage de chaque spot. Des logiciels d'extraction d'image permettent de faire le lien entre l'image du scanner et les spots du microarray (Agilent's Feature Extraction, GenePix Pro par exemple). Outre l'intensité des spots, le signal à l'extérieur des spots est aussi mesuré pour pouvoir estimer le bruit de fond, qui correspond à l'hybridation non spécifique de la cible autour d'un spot.

Lorsque deux échantillons sont hybridés sur le même microarray (Agilent Technology®), le scanner lit l'intensité du marquage à deux longueurs d'onde et permet d'obtenir une valeur d'intensité pour chaque échantillon.

2.1.1.2.3 Nettoyage et normalisation des données

Afin d'améliorer la qualité des résultats obtenus par microarray, il est important d'identifier les sources de variabilité, de les contrôler lors de la phase expérimentale et de diminuer leur effet en filtrant et normalisant les données.

Sources de variabilité des analyses microarray

Pour valider les différences d'expression des gènes entre deux conditions, il faut s'assurer que la variabilité observée entre ces deux conditions ne provient pas d'autres sources que celle que l'on veut étudier. Il est donc important d'identifier ces sources de variabilité et de mettre en place un protocole expérimental qui permet de les contrôler.

Une première source de variabilité des données de microarray peut provenir du microarray lui-même avec des artéfacts lors de sa conception (au moment du dépôt des sondes surtout pour les ADNc ou lors de la synthèse des oligonucléotides surtout si elle a lieu avant le dépôt sur la lame). Cependant, la technologie de synthèse *in situ* des sondes a permis de bien améliorer ce défaut et actuellement la préparation des lames est bien plus constante, ce qui augmente la reproductibilité des lames. Des sources de variations peuvent aussi provenir de la cible, à cause de l'échantillonnage ou de la qualité des ARN. En effet, chaque étape de la préparation des cibles (purification des ARN, transcription inverse en ADNc, amplification éventuelle par PCR, marquage, hybridation des lames) représente une source de variabilité entraînant des artéfacts (Lee 2004). Il est important de vérifier la qualité des ARN et de n'hybrider que des ARN de bonne qualité (avec des valeurs de RIN proche de 9). L'efficacité de l'hybridation dépend aussi de nombreux facteurs : température, durée de l'incubation, lavage, fluorochrome utilisé, hybridations non spécifiques, conservation des échantillons, manipulateur, ... (Schuchhardt et al. 2000; Lee 2004). Ainsi, quelques différences peuvent exister pour un même échantillon hybridé sur deux microarrays. Lee (2004) rapporte aussi que les procédures d'analyse d'images peuvent engendrer de la variabilité, c'est pourquoi il est recommandé de ne pas modifier les paramètres du scanner lors de différentes prises de mesure.

Toutefois, la reproductibilité des analyses de microarray a été étudiée par hybridation de deux triplicats sur des puces avec 65 oligonucléotides et une faible variabilité a été mise en évidence entre les tripliquats, ce qui indique une bonne reproductibilité des processus de microarray (Lee et al. 2007).

Normalisation des données

Pour diminuer l'influence des différentes sources de variabilité, il est possible de normaliser les données brutes issues de la lecture du scanner avec des logiciels (R package LIMMA *Linear Models for Microarray Data*, GeneSpring®, par exemple). La première étape consiste à filtrer les données, c'est-à-dire à éliminer des spots pour lesquels l'intensité du marquage est très faible ou pas uniforme sur le spot, le bruit de fond (l'intensité de marquage autour des spots) est élevé, ou qui sont saturés (Figure 7 A et B et Figure 8 A et B).

L'analyse des microarrays repose sur le postulat que la majeure partie des gènes ne diffère pas d'une condition à l'autre. Ainsi la distribution des intensités sur l'ensemble des microarrays au cours d'une expérience doit être similaire. Pour les hybridations en double couleur, il existe un biais lié à l'incorporation des deux fluorochromes qui ont des

encombrements stériques et des intensités de fluorescence différente. Ces données sont généralement décrites par deux variables : **A** qui représente la moyenne des log des intensités rouges (Cy5) et vertes (Cy3) et **M** qui symbolise la différence des log.

$$A = (\log_2(Cy5) + \log_2(Cy3)) / 2$$

$$M = \log_2(Cy5) - \log_2(Cy3)$$

Ces données sont représentées par des MAplot (Figure 8). Le biais entre Cy3 et Cy5 se traduit généralement par un excès de faibles valeurs pour la Cy3, ce biais est visible par le grand nombre de valeurs dans le carré inférieur gauche dans le MAplot tracé avec les données filtrées (Figure 8B). La fonction *normalizeWithinArrays* du package LIMMA (Smyth et Speed 2003; Smyth 2005) permet d'appliquer une fonction *loess* aux distributions des intensités des deux canaux pour rapprocher les distributions des canaux Cy3 et Cy5 et ainsi éliminer ce biais lié au marquage (Figure 7B et C et Figure 8B et C).

Ensuite, pour corriger tous les biais liés aux différences d'hybridation d'un microarray à un autre, la fonction *normalizeBetweenArrays* du package LIMMA (Smyth 2005) est bien adaptée. Elle permet par exemple d'imposer la même distribution empirique à tous les microarrays et tous les canaux (**method=quantile**) (Yang et Thorne 2003).

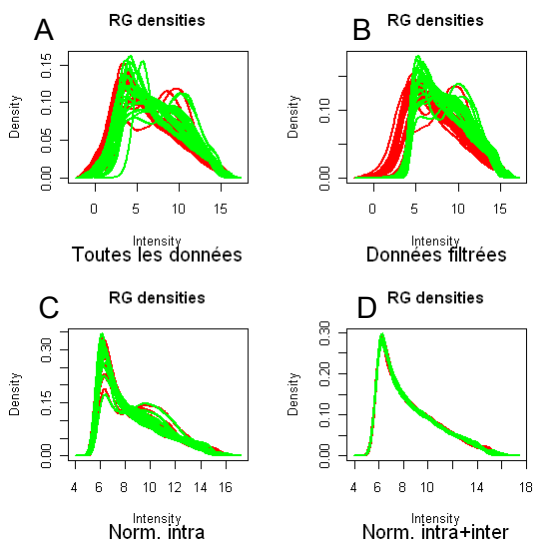


Figure 7 : Distribution des intensités de fluorescence de tous les spots d'un microarray selon la normalisation des données. Figure (A) : données brutes. Figures B : données filtrées. Figure C : données filtrées et normalisées (intra-lame loess) et Figure D : données filtrées et normalisées (intra-lame loess + inter lames A quantile sans correction du bruit de fond).

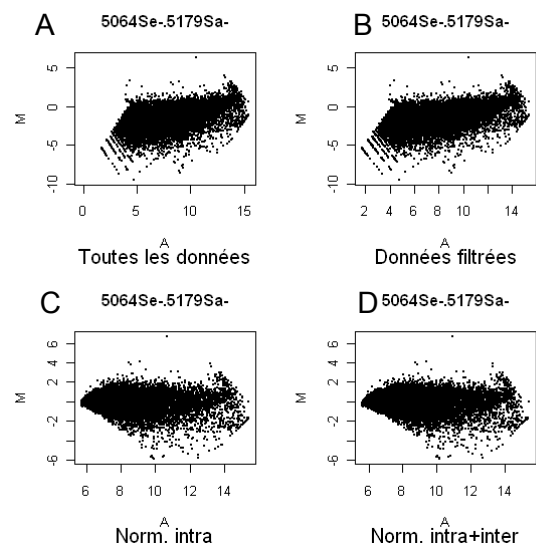


Figure 8 : MAplot des intensités de fluorescence selon la normalisation des données de microarray. Le A en abscisse correspond à la moyenne des intensités des deux fluorochromes et le M en ordonnée à la différence entre les intensités de Cy5 et de Cy3. Les lettres des figures sont explicitées dans la figure de gauche.

Reproductibilité des analyses de microarray

Nous venons de voir qu'il existe plusieurs types de microarray, que les échantillons peuvent être marqués de différentes façons, que les protocoles d'hybridation et de lecture des données sont variés, qu'il existe de nombreuses méthodes de traitement des données. Tout cela rend la reproductibilité d'une expérience difficile. Les membres du Microarray Gene Expression Data (MGED <http://www.mged.org/>) ont travaillé ensemble pour récapituler les données standards à fournir pour la publication de données de microarray dans le MIAME (Minimal Information About a Microarray Experiment, Brazma et al. 2001). Le MIAME est découpé en six grandes sections :

- 1- Le plan expérimental qui correspond à l'ensemble des hybridations ;
- 2- Le modèle du microarray avec description de chaque puce et de chaque élément (spot, contrôle) sur la puce ;
- 3- Les échantillons avec leur nature, leur extraction et leur marquage ;
- 4- Les hybridations avec les procédures et paramètres utilisés ;
- 5- Les mesures obtenues à partir des images avec les données brutes ;
- 6- Les méthodes de normalisation (description et valeurs normalisées obtenues).

Le MIAME propose que l'ensemble de ces informations soit déposé dans des bases de données publiques (Brazma et al. 2001), comme Array Express (<http://www.ebi.ac.uk/arrayexpress/>) ou GEO –Gene Expression Omnibus- (<http://www.ncbi.nlm.nih.gov/geo/>) (Ball et al. 2002). En 2009, Brazma est revenu sur les succès, les échecs et les défis du MIAME (Brazma 2009). La plupart des journaux demande actuellement que les expériences suivent le guide MIAME pour pouvoir être publiées. Plus de 10 000 études de microarray ont été déposées dans des banques de données publiques. Cependant la reproductibilité des études de microarray déjà publiées reste limitée. Ioannidis *et al.* (2009) ont ré-analysé 18 jeux de données de microarray publiées dans Nature Genetics entre 2005 et 2006. Ils n'ont réussi à reproduire que 2 analyses complètement, 6 partiellement ou avec des contradictions, et 10 n'ont pas pu l'être du tout. Ce faible nombre s'explique par des problèmes de disponibilité des données, des incohérences dues à des problèmes d'annotation et de traitements/analyses des données (Ioannidis et al. 2009). Ils concluent en disant que les efforts doivent être poursuivis, pour que toutes les informations nécessaires pour ré-analyser un jeu de données soient disponibles publiquement.

2.1.2 Etude de la variabilité biologique dans l'expression des gènes

Quelle que soit l'approche utilisée pour mesurer l'expression d'un gène, une fois que les données ont été préparées, le but est de comparer la quantité d'ARNm entre différentes conditions afin de mettre en relation l'expression de certains gènes et un phénotype d'intérêt. La mesure d'expression d'un grand nombre de gènes par rapport au nombre restreint d'échantillons biologiques pose des problèmes de dimension. Le but de l'analyse statistique est alors d'extraire une information pertinente concernant l'effet de différents facteurs sur l'état fonctionnel de la cellule ou du tissu. Il existe principalement trois stratégies applicables à l'analyse de données transcriptomiques ; i) des méthodes issues de la statistique exploratoire (ACP, positionnement multidimensionnel, classification), ii) de la modélisation (analyse de variance, modèles mixtes, tests), ou iii) de l'apprentissage (forêts aléatoires avec la recherche de gènes les plus discriminants). Après avoir testé ces méthodes sur un jeu de données, des statisticiens de Toulouse déclarent qu'il n'y a pas, a priori, de meilleure approche mais qu'il faut trouver la « bonne » stratégie associant exploration et modélisation, adaptée tant aux données qu'à l'objectif recherché (Baccini et al. 2005). Outre les différentes méthodes, il est

possible de faire varier différents paramètres ou d'adapter les analyses selon les besoins et la question biologique posée. Par exemple, l'analyse de variance (ANOVA) peut être effectuée soit gène par gène, soit par groupe de gènes ayant une variabilité homogène ou hétérogène.

Dans mon manuscrit, je ne développerai que les approches de modélisation de la recherche de gènes différentiels adaptées aux données de microarray et de qPCR. Puis je donnerai des informations pour l'interprétation des résultats de listes de gènes différentiels.

2.1.2.1 Identification de gènes différemment exprimés

Le choix de l'analyse statistique pour identifier les gènes différentiels repose en partie sur le nombre d'observation afin d'accepter ou non les hypothèses du modèle. Ainsi pour les données de microarray pour lesquelles une quarantaine d'observations étaient disponibles, j'ai utilisé l'analyse de variance (*ANALYSIS OF VARIANCE*, ANOVA), alors que pour les données de qPCR avec seulement une dizaine d'observations, j'ai privilégié un test non paramétrique.

2.1.2.1.1 Analyse de variance avec des modèles linéaires

L'ANOVA permet d'étudier le comportement d'une variable à expliquer continue en fonction d'une ou plusieurs variables explicatives.

Par exemple dans l'analyse des cellules dendritiques, j'ai appliqué le modèle suivant à chaque gène :

$$y_{ijkl} = \mu + A_j + B_k + C_l + \varepsilon_{ijkl}$$

où y_{ijkl} représente la $i^{\text{ème}}$ expression d'un gène dans la lignée j , au temps k , avec le marquage de l'échantillon l ;

A_j est l'effet de la lignée à 2 modalités : A_1 pour résistant et A_2 pour sensible ;

B_k est l'effet du temps de stimulation à 3 modalités : B_1 pour T0, B_2 pour T3 et B_3 pour T8 ;

C_l est l'effet du marquage de l'échantillon à 2 modalités : C_1 pour Cy3 et C_2 pour Cy5 ;

ε_{ijk} représente le résidu du modèle distribué selon une loi normale centrée de variance σ^2 .

L'ANOVA repose sur des hypothèses qu'il est nécessaire de vérifier : normalité des résidus, homoscedasticité des variances et indépendance des échantillons. Un microarray comporte des milliers de spots, et pour chaque spot, une analyse de variance est réalisée. Il est donc difficile de vérifier les postulats pour chaque analyse. On considère généralement que l'utilisation d'un nombre suffisant d'observations par spot (généralement $n = 30$) permet d'accepter les hypothèses. La normalité est vérifiée globalement, sur l'ensemble des spots de la puce, après transformation logarithmique (ici de base 2) des intensités brutes.

Le test de Fisher de chaque effet du modèle permet d'obtenir un niveau de signification de cet effet pour le spot étudié, en fonction du risque de première espèce α .

2.1.2.1.2 Correction sur les tests multiples

Dans les analyses de microarray, puisque une ANOVA est appliquée à chaque gène, soit plus de 10 000 analyses, le risque de déclarer une sonde différentielle alors qu'elle ne l'est pas peut s'avérer grand dans la liste finale. Il convient alors de corriger le seuil de signification pour la multiplicité des tests effectués. Il existe plusieurs méthodes, la plus

stricte, celle de Bonferroni, recalcule le risque de première espèce α en divisant α par le nombre de tests. Cette correction diminue fortement le taux de faux positifs, mais augmente le taux de faux négatifs. Généralement dans les analyses de microarray des corrections plus souples sont appliquées, la plus courante est celle de Benjamini Hochberg (Pawitan et al. 2005). Elle est basée sur les étapes suivantes (Benjamini et Hochberg 1995) : i) classement des p-values de la plus faible à la plus élevée (soit k l'ordre des tests), ii) maintien de la plus forte p-value telle quelle, et chaque p-value est multipliée par le nombre total de tests (n) divisé par son rang (k). Si cette p-value corrigée (ici, appelée q) est inférieure au risque de première espèce fixé, alors ce test est significatif.

$$q = p \times (n/k)$$

si $q < \alpha \Rightarrow$ test significatif

si $q > \alpha \Rightarrow$ test non significatif

2.1.2.1.3 Mesure des différences d'expression entre les conditions biologiques

L'analyse de variance permet d'obtenir les moyennes prédites de l'intensité d'expression d'un spot pour les différentes conditions biologiques testées, corrigées pour les autres effets du modèle. Elle permet également d'obtenir un test (test de Fisher) sur l'ensemble des contrastes pour les effets.

Lorsque les données ont été transformées par la fonction logarithmique de base 2 avant les analyses de variance, l'estimation de la moyenne correspond au logarithme du fold-change (FC). Le FC est souvent utilisé, en complément de la p-value (ou p-value corrigée par les tests multiples) pour identifier les sondes différentielles. Le seuil de FC utilisé est très variable dans les analyses de transcriptome, parfois aucune contrainte n'est appliqué sur le FC, parfois un seuil de 1,2, parfois de 1,5, 2, 5 ou 10. Biologiquement une différence de 20 ou 50% d'expression entre deux conditions (ce qui correspond à un FC de 1,2 ou 1,5) peut avoir des conséquences non négligeables sur les voies métaboliques. Il peut s'agir également d'un moyen pragmatique de réduire une grande liste de gènes différentiels aux différences les plus marquées.

2.1.2.1.4 Cas particulier des données de RT-qPCR

Pour les données de RT-qPCR, je disposais de 8 à 60 observations (12 en moyenne). Les hypothèses de normalité et d'homoscédasticité des variances requises pour l'ANOVA étaient souvent non respectées. Il était donc préférable d'utiliser un test non paramétrique qui ne fait pas d'hypothèse sur la distribution des échantillons, comme par exemple le test de Wilcoxon, aussi appelé test de Mann-Whitney.

Ce test repose sur la stratégie suivante : i) on rassemble les deux séries d'échantillons et on range toutes les valeurs par ordre croissant ; ii) on calcule la somme des rangs des X_i et des Y_i (statistiques de Wilcoxon). La statistique du test de Wilcoxon est tabulée en fonction du nombre d'échantillons et donc du nombre de permutations possibles ; iii) on peut ainsi dire si les deux distributions sont égales (acceptation de l'hypothèse nulle) ou différentes (rejet de l'hypothèse nulle). La procédure *npar1way* de SAS permet de calculer le test de Wilcoxon.

2.1.2.2 Interprétation des listes de gènes

La partie précédente a permis d'identifier une liste de gènes différentiels. Lorsque la liste comporte peu de gènes, il est possible de consulter la littérature pour rechercher des informations sur chacun des gènes et sur leurs interactions. Cependant lorsque le nombre de gènes différentiels est grand, il est alors difficile de considérer individuellement chaque gène, et l'interprétation globale de la liste de gènes est recommandée. Il existe différentes bases de données qui recueillent soit des informations bibliographiques, soit des informations de séquences ou de fonctions des gènes connus et qui permettent d'établir des relations entre les différents gènes différentiels.

2.1.2.2.1 Les annotations de la puce ovine Agilent

Avant d'interpréter une liste de gènes, il est important de disposer d'une bonne annotation de celle-ci. La puce ovine Agilent (019921) possède 15208 sondes. Chaque sonde est désignée par un nom spécifique à Agilent. D'autres annotations sont disponibles par le constructeur de la puce : 93% des sondes ont un « TargetID » qui permet d'identifier les sondes dans certaines bases de données (NCBI notamment) et seulement 10 % des sondes ont une annotation HUGO qui est une nomenclature standardisée des gènes humains et qui est très utile pour interpréter les gènes.

L'équipe de SIGENAE a développé un outil sigReannot (Casel et al. 2009) qui permet d'annoter les séquences des sondes des microarrays pour vérifier et améliorer les annotations. Les annotations de la puce ovine Agilent sont disponibles sur le site de Sigenae (<http://www.sigenae.org/> data : oligoset annotation / download annotation file : Agilent / Sheep 15K). La version 6 (juin 2010) que j'ai utilisée comportait 7 529 gènes ; la version actuelle (version 8, mars 2011) en contient 8 547. L'annotation globale de la puce est alors passée de 49,5 % à 56,2%.

Ainsi, pour chaque liste de gènes différentiels, une partie était annotée avec les données d'Agilent, une autre avec Sigenae, mais il manquait des annotations. Pour compléter la liste, j'ai utilisé la fonction BLAST de NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome, rubrique « nucleotid blast »). Je me suis référée à la base de données « Others », qui ne correspond ni à l'homme, ni à la souris et j'ai utilisé le programme de sélection qui optimisait la similarité des séquences (« blastn »). Ainsi à partir de l'annotation « Accession Number », j'ai pu obtenir le « Gene Symbol » qui correspond au nom HUGO. Grâce aux différentes approches d'annotation, seul un faible pourcentage (entre 1 et 5%) des sondes différentielles est resté non annoté.

2.1.2.2.2 Analyses de groupes de gènes

Pour analyser conjointement l'ensemble des gènes, il existe des méthodes qui permettent de regrouper ces gènes, selon leur niveau d'expression (clusters), selon leurs interactions les uns aux autres (réseaux de gènes), selon leurs grandes fonctions biologiques (analyses de Gene Ontology, GO), selon les voies dans lesquels ils sont impliqués, selon les facteurs de transcription qui peuvent les réguler. Je vais aborder ces différents points dans les paragraphes suivants.

Clusters de gènes

Une première approche pour donner du sens aux analyses transcriptomiques est de regrouper les gènes en fonction de l'évolution de leur niveau d'expression entre différentes conditions. Pour cela, on peut faire une classification hiérarchique des gènes qui permet de regrouper dans un cluster des gènes dont le niveau d'expression dans différentes conditions est corrélé. Plusieurs logiciels (R, package LIMMA, GeneSpring®) permettent de tracer ces graphiques, appelés *heatmap*. Ils représentent dans les couleurs chaudes des gènes surexprimés et dans les couleurs froides des gènes sous-exprimés (Figure 9). Dans un heatmap chaque ligne correspond à un gène différentiel et chaque colonne correspond soit à un échantillon (Figure 9A), soit à une moyenne de tous les échantillons dans une condition (Figure 9B). L'arbre des corrélations entre les différents groupes de gènes (à gauche des graphiques) permet de déterminer des sous-groupes de gènes (*clusters*) qui ont des niveaux d'expression fortement corrélés.

L'élaboration de clusters ne permet pas de donner du sens biologique à une liste de gènes, mais elle permet de regrouper des gènes qui sont surexprimés ou inhibés dans certaines conditions pour interpréter plus finement les listes de gènes.

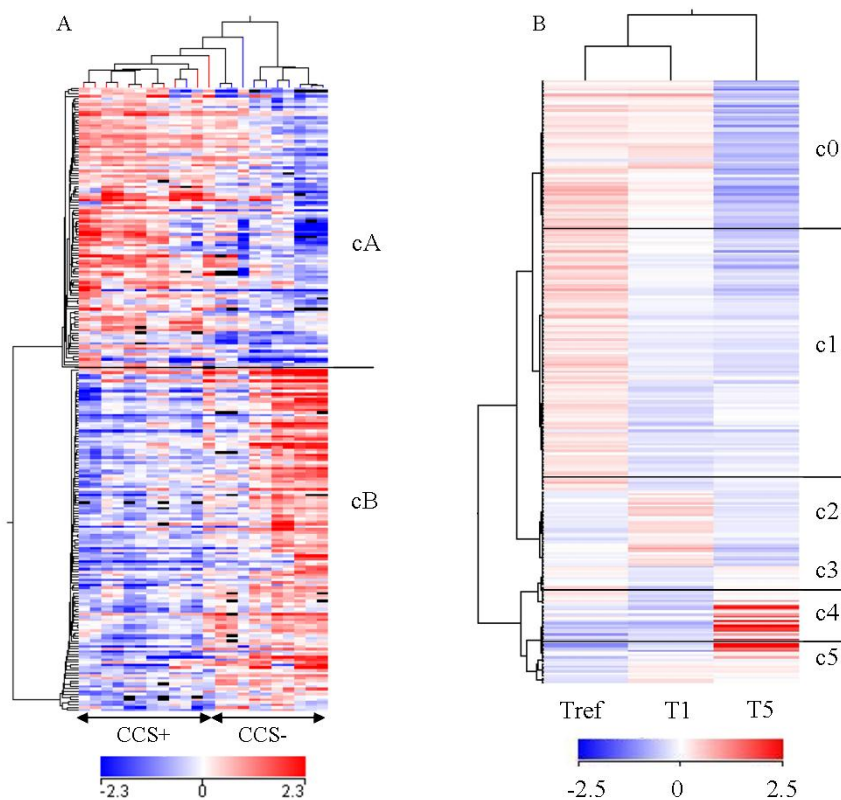


Figure 9 : Classification hiérarchique des gènes avec un heatmap. Chaque ligne correspond à un gène, chaque colonne correspond à un échantillon (figure A) ou à la moyenne de tous les échantillons à un temps donné au temps de référence (Tref), après 1 ou 5 h de stimulation (T1 et T5, respectivement) pour ce gène donné (figure B). Les arbres sur la gauche des graphiques permettent de déterminer des sous-groupes (*clusters*) de gènes qui ont des variations d'expression fortement corrélées.

Analyses des fonctions et des voies biologiques

Une autre approche pour donner du sens biologique à la liste de gènes différentiels consiste à rassembler les gènes autour de grandes fonctions biologiques déjà connues. Il existe

plusieurs bases de données dans lesquelles de nombreux gènes sont renseignés et rattachés à une ou plusieurs voies métaboliques.

Par exemple dans la base de données Gene Ontology (GO, <http://www.geneontology.org/>), les gènes sont classés dans différentes identités des trois grandes ontologies : i) les compartiments cellulaires (*cellular component, CC*), ii) les fonctions moléculaires (*molecular function, MF*), et iii) les processus biologiques (*biological process, BP*).

La base de données KEGG (KEGG, <http://www.genome.jp/kegg/>) est composée de seize bases de données principales, rangées dans différentes catégories : information systémique, information chimique et information génomique qui sont largement utilisées comme référence pour l'interprétation biologique de jeux de données générés par les analyses à haut-débit.

La base de données Ingenuity Pathway Knowledge Base (IPKB, http://www.ingenuity.com/products/pathways_knowledge.html) dérive des fonctions et interactions de gènes entre eux qui sont publiées dans la littérature. A chaque gène est attribué une fonction (par exemple « croissance cellulaire et prolifération ») et une sous-fonction (comme « formation de colonie »).

Dans chacune des bases de données Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>) et Ingenuity Pathway Knowledge Base (IPKB, http://www.ingenuity.com/products/pathways_knowledge.html), respectivement des « KEGG pathways » ou des “canonical pathways” sont prédéfinis.

Des logiciels tels que DAVID (Database for Annotation, Visualization and Integrated Discovery, version 6.7, <http://david.abcc.ncifcrf.gov/>), IPA (Ingenuity Pathway Analysis, <http://www.innatedb.com/>), ou InnateDB (Innate Data Base, <http://www.innatedb.com/>) permettent de consulter ces bases et d'établir une valeur statistique pour savoir si une fonction biologique ou une voie métabolique est assignée à cette liste de gènes différentiels par hasard ou pas.

Analyses de réseaux de gènes

Une analyse complémentaire consiste à étudier les différents liens qui existent entre les gènes différentiels deux à deux. Ces liens peuvent être de différente nature : régulation, répression, phosphorylation, oxydation ... Dans les analyses de réseau, comme le proposent IPA et InnateDB, ces informations sont répertoriées et représentées sous forme de réseau de gènes. Dans un réseau, les gènes correspondent à des points et les arêtes à des relations biologiques qui ont été publiées auparavant. D'autres gènes, qui interagissent avec un ou plusieurs gènes de la liste étudiée peuvent être ajoutés aux réseaux pour compléter l'information.

Selon le logiciel utilisé les annotations de gène à fournir sont différentes. Le logiciel IPA accepte vingt-cinq annotations différentes, dont Ensembl, EntrezGene, GenBank, HUGO Gene Symbol, RefSeq, ... tandis que le logiciel InnateDB n'en accepte que cinq : InnateDB, Ensembl, EntrezGene, RefSeq et Uniprot. Il est possible de convertir les HUGO Gene Symbol en annotation Ensembl, à partir d'un site accessible en ligne (<http://www.ensembl.org/index.html>), par l'intermédiaire du « Biomart » et de la base de données Ensembl Genes 62. Pour les analyses de réseau, IPA et InnateDB utilisent les p-values des tests de significativité et/ou les FC pour chaque gène. Ensuite, les logiciels

établissent un ou plusieurs réseaux en incorporant des gènes reliés par une ou plusieurs arêtes et calculent des scores ou des p-values pour ces réseaux selon leur vraisemblance.

Dans IPA, l'utilisateur choisit le nombre maximum de molécules à incorporer dans un réseau et le nombre maximum de réseaux générés. Puis le logiciel calcule un score en fonction du nombre de gènes dans le réseau et du nombre de gènes dans la liste en prenant en compte la p-value de chaque gène et son FC. Plus le score d'un réseau est grand, plus ce réseau apporte de l'information et plus il est fiable. Les scores peuvent atteindre 50 unités, et en général on considère qu'un réseau est informatif si son score est supérieur à 20. De plus chaque réseau est qualifié par trois grandes fonctions qui caractérisent certains gènes du réseau. Il existe différents types de réseaux. Par exemple dans le réseau 1 (Figure 10), TP53 représente un noyau, *ie* c'est un gène qui structure fortement ce réseau, car de nombreux gènes gravitent autour de lui et si on l'enlève la structure du réseau est complètement bouleversée. Ainsi, ce gène a une influence forte sur le phénomène étudié. Une légère surexpression ou inhibition de ce gène va se répercuter sur l'expression de nombreux gènes. En revanche, dans le réseau 2 (Figure 10), on ne retrouve pas cette notion de noyau. Mais on voit globalement plusieurs gènes dont l'expression varie selon les conditions et qui ont des interactions entre eux.

InnateDB permet de faire des analyses de réseaux comme IPA, mais il se concentre principalement sur l'étude des gènes de l'immunité innée.

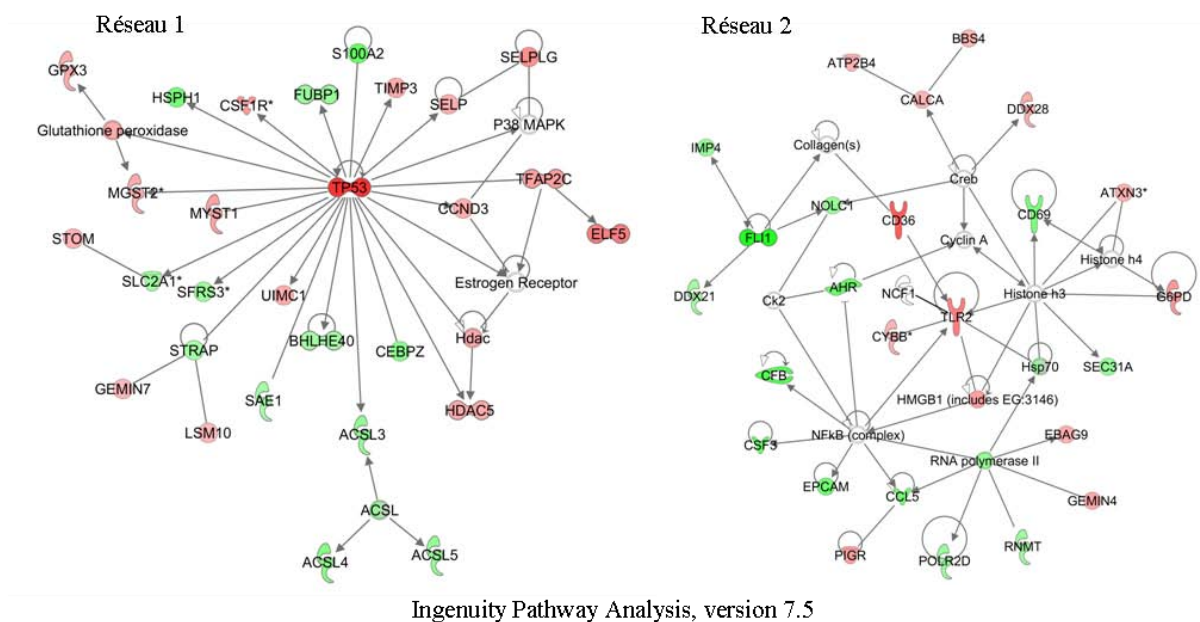


Figure 10 : Exemples de deux réseaux de gènes obtenus avec IPA. Les gènes dont l'expression est augmentée sont en rouge et les gènes dont l'expression est inhibée sont en vert.

Analyse des facteurs de transcription

Plusieurs gènes, présents dans une liste de gènes différentiels, possèdent parfois un même fragment de séquence, notamment dans la région promotrice sur laquelle peut se fixer des facteurs de transcription. Ces séquences correspondent aux sites de liaison à un facteur de transcription (« Transcription Factor Binding Site », TFBS). Si un TFBS est présent plusieurs fois dans une liste de gènes différentiels, il est possible que ces gènes soient sous la régulation de ce même facteur de transcription. Ainsi la recherche de facteurs de transcription qui

régulent plusieurs gènes différentiels est intéressante dans la perspective de recherche d'une mutation causale.

Le logiciel InnateDB permet d'identifier des facteurs de transcription qui régulent des gènes présents dans une liste d'intérêt.

2.1.2.3 Analyse d'enrichissement

En parallèle de la recherche de gènes différentiels et de l'interprétation de la liste de gènes considérée, il est possible de faire des analyses d'enrichissement pour savoir si une fonction biologique est activement représentée dans un jeu de données. Avec cette approche, on s'intéresse plus aux fonctions des gènes qu'aux gènes eux-mêmes. La méthode de calcul GSEA (Gene Set Enrichment Analysis, <http://www.broadinstitute.org/gsea/index.jsp>) permet de déterminer si dans une liste de gènes, un groupe de gènes (*gene set*) défini *a priori* a des valeurs différentes entre deux conditions. Les principales étapes de l'analyse GSEA sont les suivantes : i) calcul de la statistique t du test de student ou du FC entre deux conditions étudiées, ii) tri des gènes selon cette valeur, iii) analyse d'enrichissement en se demandant si les gènes d'une liste d'intérêt qui appartiennent aussi à *gene set* sont distribués aléatoirement. Dans la base de données GSEA, il existe déjà cinq collections de *gene sets*, qui se basent sur la localisation des gènes (C1), sur des voies métaboliques établies (C2), sur des motifs de séquences régulatrices (C3), sur des calculs (C4, basé sur des analyses de cancer) et sur les annotations GO (C5). Il est possible de définir soi-même de nouveaux jeux de gènes ; iv) calcul de la p-value ou du score du test d'enrichissement v) permutations des données et calcul des p-values ou scores, vi) comparaison des p-values obtenues avec le jeu de données et de celles obtenues après permutation et estimation de l'enrichissement pour un set de gènes.

L'analyse GSEA est principalement utilisée, lorsque très peu, voire pas du tout de gènes ont été identifiés par analyse de gènes différentiels, ou en complément de l'analyse de gènes différentiels.

2.1.3 Comment confirmer les résultats ?

Les analyses de microarray permettent d'identifier des gènes différemment exprimés entre plusieurs conditions. Bien que la préparation des échantillons, leur hybridation et l'analyse statistique des résultats soient très rigoureuses, il est généralement recommandé de valider les gènes différentiels au moins en partie par une approche différente.

Classiquement, l'expression de quelques gènes différentiels est mesurée par RT-qPCR quantitative à partir d'amorces spécifiques de ces transcrits. Ces gènes sont choisis soit car ils ont un FC élevé soit pour leurs propriétés biologiques. Si possible, les RT-qPCR sont réalisées sur des échantillons indépendants de ceux testés en microarray. Mais lorsque l'obtention de réplicats biologiques est difficile les confirmations sont établies sur les mêmes échantillons.

D'autres approches pour valider un gène différentiel sont également possibles. Par exemple, l'expression de la protéine associée à l'ARNm peut être mesurée, mais il a été montré qu'une augmentation de l'ARNm n'augmentait pas toujours celle de la protéine associée en relation avec toutes les étapes de maturation des ARNm et du décalage dans le temps. L'implication de ce gène dans une voie métabolique peut aussi être mise en évidence par inhibition ou activation de l'expression de ce gène dans un modèle cellulaire ou chez la

souris. Les études de souris *knock out* peuvent permettre de démontrer le rôle d'un gène dans un processus métabolique.

La comparaison du transcriptome des brebis de la lignée CCS+ et de celles de la lignée CCS- peut permettre d'identifier des mécanismes qui sont différemment régulés au sein de chaque lignée. Ils peuvent aussi permettre de formuler de nouvelles hypothèses sur des gènes fonctionnels candidats dont l'expression peut être associée à des formes de résistance ou de sensibilité aux IMI.

2.2 Etude du polymorphisme de l'ADN

L'étude du polymorphisme de l'ADN entre les deux lignées de brebis peut permettre à la fois de confirmer des hypothèses qui ont été formulées à partir des analyses du transcriptome et d'identifier de nouvelles régions chromosomiques associées aux CCS, voire d'émettre de nouvelles hypothèses de mutations/gènes candidats responsables des différences de sensibilité aux mammites.

2.2.1 Un outil de choix pour l'étude du polymorphisme chez l'ovine : la puce SNP 54K d'Illumina

Grâce aux avancées technologiques récentes en génétique moléculaire, il existe depuis 2009 un outil à haut débit de choix pour étudier le polymorphisme de l'ADN chez le mouton : la puce SNP ovine (BeadChip ovine 54K) développée par Illumina en collaboration avec le consortium International ovine (International Sheep Genomics Consortium, ISGC <http://www.sheephapmap.org/>). Elle contient 54 241 SNP, qui ont été détectés grâce au séquençage de 60 animaux de quinze races ovines différentes. Ces SNP ont été choisis pour être polymorphes dans la plupart des races ovines. Ils ont aussi été définis de façon à ce qu'ils soient répartis uniformément sur l'ensemble du génome ovine. L'espacement entre deux SNP est de 50.9 kbases en moyenne (42.5 kbases en médiane). Ils ont aussi été définis de façon à ce que le risque de se tromper lors de l'attribution d'un génotype à un animal soit le plus faible possible, ce qui correspond à un *call rate* supérieur à 99% pour chaque SNP de la puce lors des tests réalisés avant la commercialisation de la puce SNP ovine.

La technologie de la puce, Infinium HD assay, repose sur une amplification du génome complet sans PCR, puis une fragmentation des molécules d'ADN et une détection des allèles en deux étapes : tout d'abord une hybridation des fragments d'ADN non marqués sur les oligonucléotides de 50 mères (sélectivité), puis l'extension d'une seule base par une enzyme avec un nucléotide marqué (spécificité).

La puce SNP ovine contient douze arrays identiques ; douze échantillons peuvent ainsi être hybridés et analysés en parallèle.

2.2.2 Les méthodes d'étude du polymorphisme de l'ADN

Il existe deux approches pour découvrir des polymorphismes de l'ADN intéressants.

i) L'approche **gène candidat** qui a déjà permis d'identifier et de caractériser l'effet de variants génétiques (Andersson et Georges 2004), comme pour le gène spécifiant la caséine alpha s1 chez les caprins (composition protéique du lait) (Grosclaude et al. 1972) ou pour le gène codant pour la protéine prion chez les ruminants (résistance à la tremblante) (Hunter

1997). Mais cette approche ne permet pas d'identifier toutes les mutations (par exemple dans les zones inter-géniques) et est souvent limitée au niveau de connaissance *a priori* de la fonction biologique étudiée. ii) La **détection de QTL** nécessite des marqueurs du polymorphisme de l'ADN répartis sur tout le génome et suffisamment nombreux pour étudier, sans *a priori*, l'association entre les variations alléliques de l'ADN et les variations du caractère quantitatif. Actuellement, les études utilisent généralement les SNP comme marqueurs.

Il existe trois grandes classes de méthodes pour la détection de QTL au sens large : les analyses de liaison (*Linkage Analysis*, LA), les analyses d'association (*Linkage disequilibrium*, LD) et des méthodes qui combinent les deux méthodes précédentes (LDLA). Ces trois méthodes et leurs applications sont présentées brièvement dans les paragraphes suivants.

Les analyses de liaison

L'**analyse de liaison** (LA) se base sur la **transmission des allèles au sein d'une famille** et donc sur la notion d'apparentement entre animaux. Si on considère un père hétérozygote à un marqueur, et avec un grand nombre de descendants, on peut séparer ses descendants en deux groupes : ceux qui ont reçu l'allèle A_1 du père et ceux qui ont reçu l'allèle A_2 du père. On mesure le phénotype des animaux de chaque groupe et s'il existe une différence significative entre les phénotypes des deux groupes, on considère qu'il existe un QTL lié au marqueur pour ce caractère.

Pour utiliser cette méthode, il est important d'avoir des dispositifs d'animaux structurés de grande taille qui maximisent la probabilité que les QTL ségrégent. Il existe des **dispositifs de familles de pères** (protocole filles ou petites-filles) qui utilisent les déséquilibres de liaison intra-famille (avec respectivement deux ou trois générations) ou en croisement (back cross, F1) avec mise à la reproduction de deux populations génétiquement éloignées de façon à générer un déséquilibre de liaison aussi important que possible dans la population des descendants. Cependant pour améliorer la précision de la localisation des QTL, il faut augmenter le nombre d'événements de recombinaison survenus dans la population étudiée. Deux stratégies peuvent être établies, i) ajouter des animaux, en pratique, un grand nombre d'animaux pour avoir une recombinaison dans la zone étudiée, ii) intensifier la densité des marqueurs. Mais ces deux stratégies sont souvent très coûteuses et longues à mettre en place.

Les analyses de liaison ont été largement utilisées dans toutes les espèces de rente pour un grand nombre de caractères, et pour la résistance aux mammites chez les ruminants laitiers en particulier (Rupp et Boichard 2003; Khatkar et al. 2004; Smaragdov 2006; Ogorevc et al. 2009; Rupp et Foucras 2010). Bien qu'un grand nombre de QTL ait été détecté pour la résistance aux mammites (80 chez les bovins et 9 chez les ovins) (Rupp et Foucras 2010), la localisation précise de ces QTL n'est pas très bonne. Les intervalles de confiance sont encore grands et renferment de nombreux gènes, ce qui rend difficile la recherche de mutations causales. Jusqu'à présent, un seul QTL associé à la résistance aux mammites a été caractérisé finement (Sugimoto et al. 2006).

Les analyses d'association

Lorsqu'un grand nombre de marqueurs est disponible (des dizaines de milliers, comme dans les analyses de SNP), une alternative aux analyses de liaison est possible. Il s'agit

d'exploiter le **déséquilibre de liaison** (*Linkage disequilibrium*, LD) pour détecter des QTL **dans la population entière** et non plus au sein des familles.

Le déséquilibre de liaison correspond à une **association non aléatoire d'allèles entre deux loci**. Si on considère deux marqueurs A et B, situés sur le même chromosome. Il existe alors quatre haplotypes A_1B_1 , A_1B_2 , A_2B_1 et A_2B_2 . Si les fréquences des allèles A_1 , A_2 , B_1 et B_2 dans la population sont toutes de 0,5, alors les fréquences attendues de chacun des quatre haplotypes sont de 0,25. Toute déviation des fréquences haplotypiques de 0,25 est considérée comme un déséquilibre de liaison, c'est-à-dire que les gènes ne sont pas associés aléatoirement.

En génétique animale, les animaux sont souvent apparentés (frères-sœurs, demi-frères-sœurs). Il faut donc prendre en compte les facteurs qui influencent la stratification de la population (sélection, goulot d'étranglement, structure familiale, pedigree, effets du milieu). Ces facteurs créent une différence de fréquence allélique entre les sous-groupes de la population, ce qui entraîne l'apparition de faux positifs ou faux négatifs (Cardon et Palmer 2003, Clayton et al 2005 et Marchini et al (2004). Certains modèles des études de déséquilibre de liaison permettent de prendre en compte les liens de parenté entre les animaux.

Les analyses combinées

Traditionnellement, les études familiales (avec des pères et leurs descendants) se basent sur des études de liaison (LA) et les études de population sur des analyses d'association (LD). Mais avec l'apparition des **puces SNP à haute densité**, les intervalles entre deux marqueurs ont été fortement réduits et les études familiales peuvent être analysées par des **études de liaison (LA)** qui assurent une **robustesse** en regard de la structure de la population, tout en intégrant le **déséquilibre de liaison (LD)** pour conserver la **puissance** de l'analyse. Il s'agit des analyses **LDLA**. Ainsi, il est possible de tester s'il existe une association entre un marqueur et un QTL en prenant en compte la transmission d'allèles des parents aux descendants et en utilisant l'information du déséquilibre de liaison dans la population observée.

Les analyses présentées auparavant peuvent s'appliquer SNP par SNP et tester l'association entre un seul marqueur et un phénotype, mais il est possible de faire des tests d'association entre plusieurs marqueurs et un phénotype (analyses multi SNP) qui peuvent soit prendre en compte l'haplotype, soit ne pas le prendre en compte.

2.2.3 Vers l'identification des gènes et mutations causales

Les méthodes précédentes permettent d'identifier plusieurs SNP qui représentent un QTL dont la fréquence génotypique est associée au phénotype étudié. Pour autant, les SNP ne représentent pas la **mutation** responsable de ce QTL, mais on fait l'hypothèse que celle-ci se situe à proximité dans un intervalle de confiance. Plusieurs stratégies peuvent être déployées pour arriver à l'identification des mutations causales.

La première approche consiste à formuler des **hypothèses sur des gènes candidats** et à vérifier ces hypothèses. La région QTL identifiée précédemment est positionnée sur le génome ovin (version 2, <http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/>). Comme on fait l'hypothèse que la mutation responsable de ce QTL est située à proximité, les gènes situés sur le génome ovin dans l'intervalle de confiance du QTL sont de bons candidats positionnels. Si un gène est un bon candidat positionnel, on peut le séquencer. Selon la longueur du gène, on peut soit le séquencer entièrement (petite taille), soit ne séquencer que

ses exons après rétro-transcription de son ARNm. Si le séquençage du gène permet d'identifier deux haplotypes qui sont associés à deux phénotypes différents, alors la mutation est identifiée et l'hypothèse de gène candidat est validée. Cependant, la mutation responsable du QTL peut être située dans une zone non codante, elle ne sera donc pas identifiée par cette méthode. De plus, les intervalles de confiance sont parfois trop larges (des dizaines de milliers de paires de bases, avec des dizaines voire des centaines de gènes) pour formuler une hypothèse sur un gène candidat. Il faut alors développer une autre approche.

La **cartographie fine** de la zone d'intérêt consiste à **ajouter de nouveaux marqueurs** génétiques dans cette zone pour affiner la localisation de la mutation. Différents marqueurs peuvent être ajoutés. En premier lieu si d'autres SNP ou microsatellites sont disponibles, ils sont incorporés. Puis il est possible de séquencer quelques animaux pour essayer d'identifier de nouveaux marqueurs (souvent des SNP) et de génotyper les animaux sur ces nouveaux marqueurs. Toutefois, ce travail d'ajout de marqueurs est souvent difficile et nécessite beaucoup de temps d'expérimentation.

Actuellement, les progrès technologiques permettent de **séquencer** la zone d'intérêt de tous les animaux disponibles, soit en séquençant tout le génome, ce qui demande beaucoup de travail de bioinformatique, soit en séquençant uniquement la zone ciblée, ce qui demande beaucoup de temps de préparation des échantillons et est aussi coûteux que le séquençage complet.

Ainsi, ces stratégies de séquençage de gènes candidats, de cartographie fine ou de séquençage complet de la zone d'intérêt permettent de réduire l'intervalle de confiance dans lequel se situe la mutation responsable du QTL. Parfois, selon les recombinaisons présentes chez les animaux étudiés, elles permettent d'identifier précisément cette mutation. Mais elles sont souvent très longues à mettre en place et extrêmement coûteuses. Lorsqu'une mutation intéressante est identifiée, ses allèles peuvent être considérés dans les programmes de sélection.

Lire
la seconde partie
de la thèse