



HAL
open science

Effet pléiotrope de la mutation R96C dans le gène SOCS2 chez la brebis laitière

Claire Oget

► **To cite this version:**

Claire Oget. Effet pléiotrope de la mutation R96C dans le gène SOCS2 chez la brebis laitière. Sciences agricoles. Institut National Polytechnique de Toulouse - INPT, 2019. Français. NNT : 2019INPT0100 . tel-02788966v2

HAL Id: tel-02788966

<https://hal.inrae.fr/tel-02788966v2>

Submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

Mme CLAIRE OGET

le mercredi 6 novembre 2019

Titre :

Effet pléiotrope de la mutation R96C dans le gène SOCS2 chez la brebis
laitière

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Génétique, Physiologie et Systèmes d'Élevage (GENPHYSE)

Directeur(s) de Thèse :

MME RACHEL RUPP

MME GWENOLA TOSSER-KLOPP

Rapporteurs :

Mme CAROLE CHARLIER, UNIVERSITE DE LIEGE

Mme PASCALE LE ROY, INRA RENNES

Membre(s) du jury :

Mme GWENOLA TOSSER-KLOPP, INRA TOULOUSE, Membre

Mme RACHEL RUPP, INRA TOULOUSE, Membre

M. OLIVIER SANDRA, INRA JOUY EN JOSAS, Membre

M. SEBASTIEN FRITZ, INRA JOUY EN JOSAS, Membre

Remerciements

J'aimerais commencer par adresser mes remerciements aux personnes qui m'ont fait l'honneur d'être membres de mon jury de thèse : Pascale Le Roy et Carole Charlier qui ont accepté de relire ce manuscrit en tant que rapporteuses, et Olivier Sandra et Sébastien Fritz, en tant qu'examinateurs.

Mes remerciements les plus sincères vont ensuite à mes chères directrices de thèse, Rachel Rupp et Gwenola Tossier-Klopp. Merci à vous deux de m'avoir confié ce travail passionnant, votre "bébé SOCS", de toujours avoir su me conseiller, m'aiguiller dans les interprétations, me redonner de la motivation chaque semaine. Merci pour votre bienveillance à mon égard, je vous en serai toujours reconnaissante. Je n'oublierai jamais ces nuits blanches à collecter les échantillons avec enthousiasme, ni ces longues heures de train pour éviter la pollution de l'avion... Et que dire de votre sourire communicatif... Merci !

Mes remerciements vont ensuite à mes tuteurs de stage de fin d'étude d'école d'ingénieurs, Isabelle Palhière et Bertrand Serwin, sans qui on ne m'aurait probablement pas proposé ce projet de thèse. Merci de m'avoir transmis cette passion de la recherche et de la génétique, de votre soutien auprès des collègues, et de votre patience pour ma première expérience de publication.

Un grand merci à tous les membres de mon comité de thèse. Merci à Hervé Acloque pour ses conseils et ses encouragements. Une nouvelle fois merci à Bertrand Serwin qui me suit depuis mes débuts à l'INRA et vers qui je me suis régulièrement tournée pour ses conseils bienveillants et pertinents. En espérant que mes mails toujours à rallonge ne l'aient pas trop ennuyé. Merci à Jean-Michel Astruc qui a toujours su se rendre disponible pour discuter des données et des résultats en rapport avec la filière. Merci à Gilles Foucras, sans qui le projet finançant cette thèse aurait probablement été inexistant, et qui a accepté de me confier la première partie de ce beau projet ANR.

Merci également aux membres de l'école doctorale SEVAB qui ont su être une oreille attentive chaque année à l'avancement de ma thèse en me fournissant des conseils avisés pour la suite de ma carrière : Claude Maranges, Hervé Rémignon, et Dominique Pantalacci.

Merci à l'INPT qui a autorisé et m'a fourni le cadre administratif me permettant de réaliser cette thèse et qui m'a également donné l'opportunité de réaliser ma première expérience dans l'enseignement. J'aimerais par ailleurs remercier Pierre Maury, enseignant-chercheur à l'ENSAT avec lequel j'ai eu l'opportunité d'avoir cette première

expérience en tant que doctorante chargée d'enseignement durant deux années consécutives au cours de ma thèse. Et merci à Elana Dayoub de m'avoir épaulée la seconde année durant ces longues heures de TD/TP de Phytotechnie.

Merci aux personnes qui ont su m'aider lorsque j'avais besoin d'avis d'experts sur la sélection génomique : Andres Legarra, Christèle Robert-Granié, Céline Carillier-Jacquin, Marc Teissier. Et également concernant l'immunologie : Gilles Foucras et Cherwin Hassel.

Merci à toutes les personnes qui ont permis la bonne conduite de l'expérimentation : de GenPhySE (Laurence, Charlotte, Louise), et de l'ENVT (Gilles, Charlotte, Laurence, Marion et Christian). Un remerciement particulier à Julien Sarry qui s'est rendu disponible pour l'expérimentation, l'extraction des ARN, la préparation des librairies... et à Cédric Cabau pour son aide précieuse qui m'a permis d'avoir un aperçu de la bioinformatique (et plein de petites astuces) à travers l'alignement des séquences.

Un grand merci à l'ensemble des collègues de GenPhySE qui m'ont accueillie ces presque quatre années dans les deux bâtiments E et C, et je remercie plus particulièrement les DU et les gestionnaires, ainsi que les équipes GeSPR et GenROC qui m'ont suivie au cours de cette thèse. Merci aux doctorants qui m'ont accompagnée : Mathieu, Tiphaine, Sophie, Marc, Hung... Et mes co-bureaux Sylvie et plus particulièrement Emilie qui a supporté plusieurs années mes sautes d'humeur... Les road- et bike-trip, les run entre midi et deux... De merveilleux souvenirs !

Merci à mes proches, à mes frères, et plus particulièrement à mes parents qui m'ont toujours soutenue aussi bien moralement que financièrement malgré la distance et le "silence radio" régulier. Vous m'avez tant transmis : l'éducation, la passion pour le vélo et la musique, et bien sûr, vos allèles (déformation professionnelle), qui m'ont toujours jusqu'à présent été très favorables... Merci !

Et pour finir, j'aimerais te remercier, toi qui partages mes journées, mes nuits, mes tristesses, mes joies, bref, ma vie... mon futur époux... Encore quelques mois à patienter avant d'associer ton -notre- nom, à mes publications... Ta présence à mes côtés, et ta patience, ont été au cours de ces trois années un soutien inestimable... "les +1 contribuent au moral des troupes", comme dirait ma chère directrice de thèse !

~ Happiness only real when shared ~ Into the Wild.

Table des matières

REMERCIEMENTS	1
TABLE DES MATIERES	3
LISTE DES ABREVIATIONS	8
LISTE DES TABLEAUX.....	11
LISTE DES FIGURES	13
LISTE DES ANNEXES.....	16
INTRODUCTION GENERALE.....	19
CHAPITRE I. INTRODUCTION BIBLIOGRAPHIQUE	22
I. Gestion de mutations dans les schémas de sélection.....	22
I.1. Origine, nature et conséquences de la variabilité génétique	22
I.1.a. Définitions générales de génétique	22
I.1.b. Les différentes sources de variabilité génétique	24
I.1.c. Architecture génétique des caractères	26
I.2. Principes généraux de l'amélioration génétique des espèces d'élevage	27
I.2.a. Principe général des schémas de sélection.....	27
I.2.b. L'objectif de sélection et les index	29
I.2.c. L'évaluation génétique/génomique.....	32
I.3. Inventaire et prise en compte dans les schémas de sélection des gènes majeurs et mutations causales en ruminants	33
I.3.a. Synthèse des principales mutations causales de gènes majeurs dans les filières laitières et viande chez les ruminants	34
I.3.b. Quelques exemples de prise en compte de mutations dans les schémas de sélection.....	39
I.3.c. Cas particulier de la mutation R96C du gène <i>SOCS2</i>	41
II. SOCS-2, une protéine impliquée dans de nombreuses fonctions biologiques	44

II.1.	La voie de transmission du signal cellulaire JAK/STAT.....	44
II.1.a.	Principe général de cette voie de transmission.....	45
II.1.b.	Les protéines JAK.....	46
II.1.c.	Les protéines STAT.....	47
II.1.d.	Autres molécules intervenant dans la voie de transmission JAK/STAT.....	48
II.2.	Les protéines SOCS.....	51
II.2.a.	Structure caractéristique et fonctions des protéines SOCS.....	51
II.2.b.	Cas particulier de SOCS-2.....	54
II.3.	Les fonctions biologiques régulées par SOCS-2.....	56
II.3.a.	Reproduction.....	56
II.3.b.	Croissance et production.....	57
II.3.c.	Santé et réponse immunitaire.....	57
III.	Les études génétiques et génomiques sur les mammites ovines.....	59
III.1.	Objectifs et démarches.....	59
III.2.	Résumé de l'article de synthèse bibliographique.....	59
III.2.a.	Synthèse.....	59
III.2.b.	Introduction.....	60
III.2.c.	Les paramètres génétiques des mammites ovines.....	60
III.2.d.	Les bases de la génomique pour la résistance aux mammites.....	60
III.2.e.	L'élevage pour l'amélioration de la résistance aux mammites.....	61
III.3.	Article de synthèse bibliographique (Oget <i>et al.</i> , 2019c).....	61
IV.	Conclusions du Chapitre I. et questions scientifiques soulevées.....	72

CHAPITRE II. ÉTUDE PROSPECTIVE DE LA PRISE EN COMPTE DE LA MUTATION *SOCS2* DANS LE CONTEXTE DE LA SÉLECTION GÉNOMIQUE..... 74

I.	Objectifs et démarches.....	74
II.	Résumé de l'article scientifique.....	75
II.1.	Introduction.....	75
II.2.	Méthodes.....	75

II.3. Résultats	75
II.4. Conclusion	76
II.5. Correctif sans conséquence pour l'article (Oget <i>et al.</i> , 2019b).....	76
III. Article scientifique (Oget <i>et al.</i> , 2019b).....	76
IV. Conclusions du Chapitre II.....	91
CHAPITRE III. ASSOCIATION DU GENOTYPE <i>SOCS2</i> AUX MAMMITES ET AUX CARACTERES DE PRODUCTION DANS DES POPULATIONS OVINES EXPERIMENTALES ET COMMERCIALES	92
I. Objectifs et démarches	92
II. Effet de la mutation <i>SOCS2</i> sur les caractères de production et le statut infectieux (Oget <i>et al.</i> , 2019a)	94
II.1. Introduction.....	94
II.2. Résultats et discussion	94
II.3. Conclusion	94
III. Communication courte (Oget <i>et al.</i> , 2019a).....	95
IV. Analyses complémentaires	103
IV.1. Effet de la mutation <i>SOCS2</i> sur la reproduction femelle	103
IV.1.a. Matériels et méthodes.....	103
IV.1.b. Résultats et discussion	105
IV.1.c. Conclusion.....	106
IV.2. Effet de la mutation <i>SOCS2</i> sur la croissance et la morphologie fonctionnelle du mâle	106
IV.2.a. Matériels et méthodes.....	106
IV.2.b. Résultats et discussion	110
IV.2.c. Conclusion.....	115
V. Conclusions du Chapitre III.	116
CHAPITRE IV. REPONSE A UNE INFECTION EXPERIMENTALE CHEZ DES BREBIS PORTANT LES DIFFERENTS GENOTYPES <i>SOCS2</i>	117
I. Objectifs et démarches	117

II.	Matériel et méthodes	118
II.1.	Plan expérimental.....	118
II.1.a.	Choix des brebis.....	118
II.1.b.	Protocole expérimental	120
II.2.	Modalités de collecte et de traitement des échantillons.....	122
II.2.a.	Suivi clinique	122
II.2.b.	Prélèvements de lait.....	125
II.2.c.	Prélèvements de sang.....	126
II.2.d.	Mesures d'autopsies	129
II.3.	Traitement bioinformatique et analyses statistiques	129
II.3.a.	Analyses des phénotypes	129
II.3.b.	Analyse des données d'expression génique par RNA-seq	132
II.3.c.	Intégration de l'ensemble des données	134
III.	Résultats et discussion.....	136
III.1.	Analyses des phénotypes	136
III.1.a.	Choix des brebis et des temps analysés d'après le profilage des SCS 136	
III.1.b.	Des phénotypes révélateurs d'un contrôle de la réaction immunitaire différent entre les sauvages et les mutants.....	137
III.1.c.	Confirmation de l'effet de <i>SOCS2</i> sur le gabarit des individus... 145	
III.2.	Analyses de l'expression génique des cellules sanguines par RNA-seq . 146	
III.2.a.	Profilage de la réponse des cellules sanguines à l'infection par <i>S. aureus</i> 147	
III.2.b.	Une régulation de la réponse immunitaire différente entre les sauvages et les mutants	149
III.3.	Intégration de l'ensemble des données	152
III.3.a.	A l'échelle du protocole, des différences fortes entre temps et plus faibles entre génotypes.....	152
III.3.b.	Des gènes associés à la différence de régulation liée à <i>SOCS2</i> .. 154	
IV.	Conclusions du Chapitre IV.	157

CHAPITRE V. DISCUSSION GENERALE ET PERSPECTIVES.....	158
I. Un effet de <i>SOCS2</i> sur la reproduction ?.....	158
II. Vers une sélection avec prise en compte de la mutation ?	160
III. Les GWAS à partir des modèles de prédiction ?.....	161
IV. Les polymorphismes de <i>SOCS2</i>	163
V. Effet de la mutation <i>SOCS2</i> sur la réponse immunitaire.....	163
VI. Démonstration de la pléiotropie du gène <i>SOCS2</i> par les études génétiques	165
REFERENCES.....	166
PRODUCTIONS SCIENTIFIQUES.....	190
I. Publications	190
II. Communications orales	190
III. Posters	190
ANNEXES	191
RESUME GRAND PUBLIC	241
GENERAL PUBLIC ABSTRACT	241
RESUMES	242

Liste des abréviations

A	Adénine
Aa	Acide aminé
ADN	Acide DésoxyriboNucléique
ALK	Activin-Like Kinase (kinase semblable à l'activine)
ANR	Agence Nationale de la Recherche
ARN	Acide RiboNucléique
BCO	Beta-Carotene Oxygenase
BLUP	Best Linear Unbiased Predictor (meilleure prédiction linéaire non biaisée)
BMP	Bone Morphogenetic Protein (protéine morphogénétique osseuse)
BMPR	Bone Morphogenetic Protein Receptor (récepteur de protéines morphogénétiques osseuses)
BoHV	Bovine herpesvirus 5 (herpès-virus bovin)
C	Cytosine
CAD	Carbamoyl-phosphate synthetase 2, Aspartate transcarbamylase, and Dihydroorotase
CCS	Comptage de Cellules Somatiques
CIS	Cytokine-Inducible SH2 domain-containing protein (protéine contenant un domaine SH2 inductible aux cytokines)
CMT	California Mastitis Test, California Mastitis Test (test de mammite de Californie)
CNTF	Ciliary Neurotrophic Factor (facteur neurotrophique ciliaire)
CSN	Casein (caséine)
CTSCCV	Centre Technique de la Salaison, de la Charcuterie et des Conserves de Viandes
DA	Discriminant Analysis (analyse discriminante)
DBD	Dna Binding Domain (domaine de liaison à l'ADN)
DGAT	DiacylGlycerol o-AcylTransferase
EBV	Estimated Breeding Value (valeur génétique estimée)
EDTA	acide ÉthylèneDiamineTétraAcétique
EGF	Epidermal Growth Factor (facteur de croissance épidermique)
ENVT	École Nationale Vétérinaire de Toulouse
FERM	Four-point-one, Ezrin, Radixin, Moesin (domaine homologue à la protéine 4.1)
G	Guanine
G-CSF	Granulocyte-Colony Stimulating Factor (facteur de stimulation des colonies de granulocytes)
GDF	Growth Differentiation Factor (facteur de différenciation de la croissance)

GenPhySE	<i>G</i> énétique, <i>P</i> hysiologie et <i>S</i> ystèmes d'Élevage
GH	<i>G</i> rowth <i>H</i> ormone (hormone de croissance)
GM-CSF	<i>G</i> ranulocyte <i>M</i> acrophage- <i>C</i> olony- <i>S</i> timulating <i>F</i> actor (facteur de stimulation des colonies de granulocytes et de macrophages)
GWAS	<i>G</i> enome- <i>W</i> ide <i>A</i> ssociation <i>S</i> tudy (étude d'association pangénomique)
IA	<i>I</i> nsémination <i>A</i> rtificielle
Idele	<i>I</i> nstitut de l'Élevage
IFIP	<i>I</i> nstitut du porc
IFN	<i>I</i> nter <i>F</i> ero <i>N</i> (interféron)
IGF	<i>I</i> nsulin- <i>l</i> ike <i>G</i> rowth <i>F</i> actor (facteur de croissance ressemblant à l'insuline)
IHAP	<i>I</i> nteractions <i>H</i> ôtes- <i>A</i> gents <i>P</i> athogènes
IL	<i>I</i> nter <i>L</i> eukin (interleukine)
IPG	<i>I</i> dentification <i>P</i> érenne <i>G</i> énéralisée
ISOL	<i>I</i> ndex <i>S</i> ynthétique <i>O</i> vin <i>L</i> ait
ITEB	<i>I</i> nstitut <i>T</i> echnique de l'Élevage <i>B</i> ovin
ITOVIC	<i>I</i> nstitut <i>T</i> echnique de l'élevage <i>O</i> VIn et <i>C</i> aprin
ITP	<i>I</i> nstitut <i>T</i> echnique du <i>P</i> orc
JAB	<i>J</i> ak <i>B</i> inding <i>p</i> rotein (protéine se liant à JAK)
JAK	<i>J</i> anus <i>K</i> inase
Kb	<i>K</i> ilobase
KIR	<i>K</i> inase <i>I</i> nhibitory <i>R</i> egion (région inhibitrice de kinase)
LIF	<i>L</i> eukemia <i>i</i> nhibitory <i>f</i> actor (facteur inhibiteur de la leucémie)
Mb	<i>M</i> égabase
MC1R	<i>M</i> elano <i>C</i> ortin <i>1</i> <i>R</i> eceptor (récepteur de la mélanocortine 1)
MHC	<i>M</i> ajor <i>H</i> istocompatibility <i>C</i> omplex (complexe majeur d'histocompatibilité)
MIR	<i>M</i> oyen <i>I</i> nfra <i>R</i> ouge
MN	<i>M</i> onte <i>N</i> aturelle
NH2	<i>a</i> mino <i>t</i> erminal <i>d</i> omain (domaine N-terminus)
OAR	<i>O</i> vis <i>A</i> Ries
OS	<i>O</i> rganisme de <i>S</i> élection
OSM	<i>O</i> nco <i>S</i> tatin- <i>M</i> (oncostatine M)
PBMC	<i>P</i> eripheral <i>B</i> lood <i>M</i> ononuclear <i>C</i> ell (cellules mononuclées périphériques sanguines)
PCA	<i>P</i> rincipal <i>C</i> omponent <i>A</i> nalysis (analyse en composantes principales)
PIAS	<i>P</i> rotein <i>I</i> nhibitors of <i>A</i> ctivated <i>S</i> tat (protéines inhibitrices des STAT actives)
PIS	<i>P</i> olled <i>I</i> ntersex <i>S</i> yndrome (syndrome d'intersexualité)
PITX3	<i>P</i> aired <i>L</i> ike <i>H</i> omeodomain

PLS *Partial Least Square (moindres carrés partiels)*
 PMN *PolyMorphonuclear Neutrophil (neutrophile polymorphonucléaire)*
 PRL..... *PRoLactin (prolactine)*
 PRNP..... *PRioN Protein (protéine prion)*
 PTP *Protein Tyrosine Phosphatase (protéine tyrosine phosphatase)*
 qPCR*quantitative Polymerase Chain Reaction (réaction en chaîne par polymérase quantitative)*
 QTL..... *Quantitative Trait Locus - locus de caractère quantitatif*
 RBP *RNA-binding protein (protéine de liaison à l'ARN)*
 REIDSOCS*Robustness, Efficiency, Inflammation and Diseases under the control of SOCS2 (robustesse, efficacité, inflammation et maladies sous le contrôle de SOCS2)*
 SCC *Somatic Cell Count (comptage de cellules somatiques)*
 SCS..... *Somatic Cell Score (score de cellules somatiques)*
 SH2..... *Src Homology 2 (domaine 2 homologue à la protéine Src)*
 SNAA *Syndicat National des Aviculteurs Agréés*
 SNP..... *Single Nucleotide Polymorphism (polymorphisme nucléotidique)*
 SOCS..... *Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines)*
 SSI..... *Stat-induced Stat Inhibitor (inhibiteur de STAT induit par STAT)*
 STAM. *Signal-Transducing Adapter Molecule (molécule adaptateur à transduction de signal)*
 STAT*Signal Transducer and Activator of Transcription (transducteur de signal et activateur de transcription)*
 SUMO *Small Ubiquitin-Related MOdifier (petit modificateur lié à l'ubiquitine)*
 SYSAAF..... *Syndicat des Sélectionneurs Avicoles et Aquacoles*
 T *Thymine*
 TAD..... *Transcriptional Activation Domain (domaine à activité transcriptionnelle)*
 TMEM..... *Transmembrane Protein (protéine transmembranaire)*
 TMM *Trimmed Mean of M-values (moyenne ajustée des valeurs M)*
 TNF *Tumor Necrosis Factor (facteur de nécrose tumorale)*
 Tyk2 *Tyrosine kinase 2*
 U *Uracile*
 UFC..... *Unité Formant Colonie*
 VS..... *Versus*
 WT..... *Wild-Type (type sauvage)*

Liste des tableaux

Tableau 1 Principales mutations causales dans des gènes majeurs de caractères d'intérêt des filières laitières et allaitantes chez les ruminants	35
Tableau 2 Principales molécules de signalisation connues dans la littérature pour être inhibées par les protéines SOCS.....	53
Tableau 3 Description des saillies en première et seconde intentions des femelles de La Fage sur les campagnes de 2012 à 2019	103
Tableau 4 Répartition des effectifs de saillies selon les génotypes SOCS2.....	104
Tableau 5 Description des phénotypes (morphologie, croissance) des mâles génotypés SOCS2 : effectifs et statistiques descriptives	109
Tableau 6 Résultats des modèles linéaires mixtes effectués sur les mesures de croissance des mâles des stations	113
Tableau 7 Résultats des modèles linéaires effectués sur les données de pointages des mâles des stations.....	114
Tableau 8 Description des brebis initialement choisies (toutes les lignes) et retenues (lignes non grisées) pour les infections expérimentales.....	119
Tableau 9 Critères d'évaluation du score clinique général sur une échelle de 0 à 14	123
Tableau 10 Critères d'évaluation des scores cliniques mammaires sur une échelle de de 0 à 12	124
Tableau 11 Effectif disponible par temps pour les analyses RNA-seq de l'infection expérimentale	127
Tableau 12 Récapitulatif des données recueillies et générées au cours de l'infection expérimentale et des analyses des variables déduites	130

Tableau 13 Statistiques descriptives (moyennes et écarts-types) et résultat des tests de comparaison de moyennes sur les variables générées au cours des autopsies de l'infection expérimentale	146
Tableau 14 Nombre de gènes différentiels entre les temps de l'infection expérimentale (analyse EdgeR, p-valeur ajustée : 0,05)	147
Tableau 15 Principales voies canoniques (ratio < 0,05 et q-valeur < 0,05) parmi les gènes à expression différentielle entre les C/C et les T/T à T56 de l'infection expérimentale d'après le logiciel Ingenuity Pathway Analysis	151

Liste des figures

Figure 1 Illustration schématique de l'information génétique dans les cellules eucaryotes (Vignal, 2011)	23
Figure 2 Les différents types de mutations ponctuelles et leur conséquence sur la séquence protéique, quelques exemples concrets	25
Figure 3 Représentation schématique de l'organisation des programmes de sélection français dans les différentes filières d'animaux de rente	29
Figure 4 Caractérisation bioinformatique de la mutation R96C dans le gène ovin SOCS2 (Rupp et al., 2015).....	43
Figure 5 Principe général de la voie de transmission du signal JAK/STAT (Vignais, 1997)..	45
Figure 6 Organisation structurale commune aux différentes protéines de chacune des familles JAK et STAT (Schindler & Plumlee, 2008)	47
Figure 7 Combinaisons connues des protéines JAK intervenant dans la transmission du signal des principales molécules utilisant la voie JAK/STAT (Murray, 2007)	47
Figure 8 Les nombreux niveaux de régulation de la voie JAK/STAT (Villarino et al., 2017)	50
Figure 9 Structure et fonction de la famille des protéines SOCS dans la voie de transmission JAK/STAT (Yoshimura et al., 2018)	52
Figure 10 Description schématique des sept phénotypes de morphologie fonctionnelle du mâle, évalués dans les centres d'élevage (Jean-Michel Astruc, communication personnelle).....	107
Figure 11 Protocole des infections expérimentales réalisée à l'ENVT en janvier-février 2018	121
Figure 12 Evolution des scores de cellules somatiques dans le lait de l'ensemble des seize brebis infectées au cours de l'infection expérimentale.....	136

Figure 13 Évolution des lsmeans intra-génotype SOCS2 de la température du rumen imputée au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10).....	138
Figure 14 Évolution des lsmeans intra-génotype SOCS2 du score clinique mammaire de l'hémi-mamelle inoculée au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10).....	139
Figure 15 Évolution des lsmeans intra-génotype SOCS2 du SCS du lait de l'hémi-mamelle inoculée au cours de l'infection expérimentale (* : p-valeur < 0,10).....	140
Figure 16 Évolution des lsmeans intra-génotype SOCS2 de la quantité de lymphocytes dans le lait de l'hémi-mamelle inoculée au cours de l'infection expérimentale (* : p-valeur < 0,10)	141
Figure 17 Évolution des lsmeans intra-génotype SOCS2 de la quantité de lymphocytes T dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10).....	142
Figure 18 Évolution des lsmeans intra-génotype SOCS2 du ratio des quantités de lymphocytes T CD4 et CD8 dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05).....	143
Figure 19 Évolution des lsmeans intra-génotype SOCS2 du ratio des quantités de lymphocytes NK dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05)	143
Figure 20 Évolution des lsmeans intra-génotype SOCS2 de la quantité de quelques composés solubles (IL-8, CCL-2, CCL-3, haptoglobine) dans le sang au cours de l'infection expérimentale (*** : p-valeur < 0,01 ; ** : p-valeur < 0,05 ; * : p-valeur < 0,10).....	144
Figure 21 Répartition des gènes différentiels entre les trois analyses (TOUT = tous génotypes, C/C = intra-génotype C/C, et T/T = intra-génotype T/T).....	148
Figure 22 PCA sur le bloc des données de comptages de gènes à tous les temps du transcriptome de l'infection expérimentale, avec (A) et sans (B) prise en compte de l'effet animal (option multilevel)	153
Figure 23 PLS-DA avec comme facteurs discriminants le temps de contrôle (A) et le génotype SOCS2 (B) sur le bloc des données de comptages de gènes à tous les temps du transcriptome de l'infection expérimentale.....	153

Figure 24 Cercle des corrélations entre les variables expliquant le plus l'axe 1 de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » pour chaque bloc : comptages de gènes (en bleu), données cliniques (en vert), données du lait (en rouge) et données du sang (en orange) à T56 de l'infection expérimentale 156

Liste des annexes

Annexe 1 Tableaux additionnels (Table S1) de l'article de synthèse bibliographique (Chapitre I.). Ce tableau présente les calculs effectués pour obtenir une moyenne pondérée de caractères de résistance aux mammites à partir des paramètres génétiques estimés dans la littérature (héritabilités pour le premier tableau et corrélations génétiques avec le caractère de production laitière pour le second)	191
Annexe 2 Figure additionnelle (Figure S1) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique (Chapitre II.) : « Visualization of linkage disequilibrium measured as squared correlation coefficient (r^2) according to distance between markers on the 50K ovine SNP chip. »	193
Annexe 3 Figure additionnelle (Figure S2) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique (Chapitre II.) : « Visualization of linkage disequilibrium ($r^2 \times 100$) between the 40 markers closest to the SOCS2 point mutation (rs868996547, in the middle). »	194
Annexe 4 Figure additionnelle (Figure S3) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique (Chapitre II.) : « Components estimations according to the different models. One-trait methods correspond to equations (1) and (2) and two-traits methods to equations (3) and (4). »	195
Annexe 5 Figure additionnelle (Figure S4) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique (Chapitre II.) : « Manhattan plots of estimated SNP effects using the best WssGBLUP approach for each phenotype (second iteration). On the left are presented analysis without the SOCS2 genotype among the markers and on the right, with the SOCS2 genotype (green point). »	196
Annexe 6 Figure additionnelle (Figure S4) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique (Chapitre II.) : « Manhattan plots of estimated variance explained by 20 adjacent SNPs using the best WssGBLUP approach for each phenotype (second iteration). The horizontal red line represents the threshold of 1% adopted in this study. On the left are presented the analyses without the SOCS2 genotype among the markers and on the right, with the SOCS2 genotype. »	201

Annexe 7 Récapitulatif de la disponibilité des données générées au cours de l'infection expérimentale (Chapitre IV.).....	206
Annexe 8 Caractérisation des lymphocytes du sang avec le logiciel FlowJo® (Sarah Walachowski, Chapitre IV.).....	207
Annexe 9 Pipeline du traitement bioinformatique des données de séquence d'ARN de l'infection expérimentale (Chapitre IV.).....	208
Annexe 10 Résultats des modèles mixtes linéaires effectués sur les phénotypes de l'infection expérimentale (Chapitre IV.).....	209
Annexe 11 Quantités moyennes par génotype SOCS2 des populations cellulaires dans le sang au cours de l'infection expérimentale (Chapitre IV.)	218
Annexe 12 Diagrammes de Venn représentant les gènes différentiels en commun ou indépendants entre les analyses RNA-seq mettant en lumière la réponse immunitaire (différences entre les temps, analyses tous génotypes SOCS2 -TOUT- ou intra-génotype SOCS2 -CC et TT-) entre T0 et les autres temps (quand le nombre de gènes est différent de 0 ou 1) (Chapitre IV.).....	219
Annexe 13 Principales (Top 10) voies canoniques parmi les gènes à expression différentielle entre quelques temps de l'infection expérimentale (Chapitre IV.)	220
Annexe 14 Gènes différentiellement exprimés entre les C/C et les T/T aux différents temps de l'infection expérimentale (Chapitre IV.).....	221
Annexe 15 Sparse multilevel multi-bloc PLS-DA avec comme facteur discriminant le « temps de contrôle » sur les blocs des données de comptages de gènes (A), des données cliniques (B), et des données lait-sang (C) à tous les temps du transcriptome de l'infection expérimentale (Chapitre IV.)	226
Annexe 16 Sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » sur les blocs des données de comptages de gènes (A), des données cliniques (B), et des données lait-sang (C) à tous les temps du transcriptome de l'infection expérimentale (Chapitre IV.)	227

Annexe 17 PLS-DA avec comme facteur discriminant le « génotype SOCS2 » sur les blocs des données de comptages de gènes (A, avec option « sparse »), cliniques (B), de lait (C), et de sang (D) à T56 (Chapitre IV.)	228
Annexe 18 Graphique des individus de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » dans les blocs de données de comptages de gènes (A), de données cliniques (B), de données du lait (C) et de données du sang (D) à T56 de l'infection expérimentale (Chapitre IV.).....	229
Annexe 19 Variables expliquant le plus l'axe 1 de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » dans les blocs de données de comptages de gènes (A), de données cliniques (B), de données du lait (C) et de données du sang (D) à T56 de l'infection expérimentale (Chapitre IV.).....	230
Annexe 20 Résumé de la communication orale présentée à la 27 ^{ème} Conférence Animal & Plant Genome (PAG) du 12 au 16 janvier 2019 (San Diego, ÉTATS-UNIS)	231
Annexe 21 Poster présenté au 20 ^{ème} séminaire des doctorants du Département de Génétique Animale (INRA) les 09-10 mai 2017 (Rennes, FRANCE).....	232
Annexe 22 Poster et papier court associé, présentés au congrès mondial de génétique animale appliquée aux animaux d'élevage (WCGALP) du 11 au 16 février 2018 (Auckland, NOUVELLE-ZÉLANDE).....	233

Introduction générale

Les schémas de sélection des animaux d'élevage permettent l'amélioration des espèces en faveur de caractères d'intérêt pour l'Homme tels que la production, la résistance aux maladies, la capacité de reproduction, ou encore l'adaptabilité aux conditions environnementales. L'existence d'une variabilité génétique entre les individus au sein d'une même espèce, et notamment des mutations, événements naturels modifiant le génome des individus d'une génération à une autre, est ce qui permet de réaliser cette sélection. Elle est réalisée sous l'hypothèse que les caractères d'intérêt sont déterminés par de nombreux gènes à effets faibles (déterminisme polygénique infinitésimal) (Fisher, 1918) dont la forme allélique peut être transmise à la descendance et dont la connaissance n'est pas nécessaire pour l'évaluation des individus. Cependant, cette hypothèse ne reflète pas toujours la réalité. En effet, certains caractères peuvent être déterminés par des gènes à effets forts (gènes majeurs) en plus de la partie polygénique. Les nouvelles technologies de séquençage développées ces 30 dernières années ont permis d'acquérir de nouvelles connaissances quant au déterminisme génétique des caractères et notamment d'avoir accès au génome, aux gènes majeurs, ainsi qu'aux mutations présentes dans ces gènes majeurs, pouvant donc avoir de grandes conséquences (favorables ou défavorables) sur les caractères. L'utilisation d'une telle connaissance est potentiellement plus précise et efficace en sélection.

Cette thèse s'intéresse à une mutation présente dans un gène pléiotrope, c'est-à-dire qui a un effet sur plusieurs caractères indépendants, le gène *SOCS2* (*Suppressor Of Cytokine Signaling 2* - suppresseur de signalisation des cytokines 2), chez la brebis laitière. Cette mutation a été découverte grâce à l'étude des bases génétique et immunologique de la résistance aux mammites (inflammation de la mamelle), un sujet de recherche majeur au cours de ces quinze dernières années pour les unités mixtes de recherche partenaires de l'INRA (Institut National de la Recherche Agronomique) GenPhySE (Génétique, Physiologie et Systèmes d'Élevage) et IHAP (Interactions Hôtes-Agents Pathogènes). Pour l'industrie laitière, les mammites représentent un poids économique important en raison de l'altération de la qualité du lait et de l'augmentation du coût de production. En effet, en petits ruminants laitiers, les pertes totales annuelles liées aux mammites ont été estimées à environ 60 millions d'euros par an en Union Européenne (Rupp & Foucras, 2010). Ainsi, en plus des mesures d'hygiène mises en œuvre, la sélection en faveur d'une meilleure résistance aux mammites est aujourd'hui considérée et implémentée dans les programmes d'amélioration de plusieurs races laitières bovines, ovines et caprines.

La mutation du gène *SOCS2* à l'étude dans cette thèse a été identifiée dans ce cadre de recherches, à partir d'études d'association pangénomique sur le comptage de cellules somatiques (CCS) dans le lait, un caractère lié aux mammites (Rupp *et al.*, 2015). Les individus porteurs de la mutation avaient une inflammation mammaire augmentée par rapport aux individus non porteurs (effet négatif sur la santé), cette mutation expliquant 12% de la variance génétique du caractère dans une population où elle était présente à une fréquence de 21,7%. Il a été démontré de plus que la taille, le poids, et la production laitière étaient également significativement augmentés chez les brebis porteuses de l'allèle mutant *SOCS2* en comparaison au phénotype sauvage (+24% pour la taille, +18% pour le poids, et +4,4% pour le lait entre les homozygotes mutés et non mutés). Cette mutation est située dans la séquence génomique qui spécifie le domaine de liaison très conservé de la protéine SOCS-2 et entraîne la perte de reconnaissance du ligand par la protéine, comme l'a montré un test fonctionnel. La protéine SOCS-2 est connue pour être impliquée dans une voie majeure de signalisation chez les Mammifères, la voie JAK (*Janus Kinase*) / STAT (*Signal Transducer and Activator of Transcription* - transducteur de signal et activateur de transcription), qui peut réguler un large spectre de cytokines et de facteurs de croissance. L'hypothèse des auteurs concernant la découverte de cette mutation est alors que ce rôle de régulateur serait à l'origine du caractère pléiotrope du gène *SOCS2* qui peut avoir un effet sur de nombreuses fonctions biologiques (santé, croissance et production). Ces résultats ont ainsi mis en lumière le rôle majeur du gène *SOCS2* comme potentiel facteur de compromis entre la réponse inflammatoire et la production laitière chez la brebis laitière, deux caractères de grand intérêt en sélection.

Ce projet de thèse s'inscrit dans le cadre des travaux en cours sur l'étude du rôle du gène *SOCS2* dans la prédisposition aux infections bactériennes, et de son effet pléiotrope à la fois sur la santé et sur les caractères de production (deux caractères sélectionnés), au sein du projet ANR (Agence Nationale de la Recherche) « REIDSOCS » (*Robustness, Efficiency, Inflammation and Diseases under the control of SOCS-2* - robustesse, efficacité, inflammation et maladies sous le contrôle de SOCS-2, [ANR-16-CE20-0010](#), 2016-2020).

Le premier objectif était de déterminer les moyens disponibles pour la prise en compte de cette mutation pléiotrope dans le contexte de la sélection génomique, mise en place aujourd'hui en brebis Lacaune sur plusieurs caractères (production, santé, et morphologie mammaire). Pour y répondre, j'ai testé différents modèles d'évaluation génétique et génomique sur les neuf caractères en sélection. Un deuxième objectif de cette étude était de confirmer, valider et découvrir l'association de cette mutation aux différents caractères d'intérêt pour la filière ovine, au-delà des seuls caractères pris en compte en sélection. Afin de répondre à cet objectif, de

nombreux phénotypes originaux étaient disponibles dans des populations ovines expérimentales et commerciales, et j'ai pu réaliser des études d'associations sur différents caractères en lien avec la santé, la production, la croissance et la reproduction. Enfin, un troisième et dernier objectif était d'identifier les mécanismes moléculaires biologiques sous-jacents de la réponse à l'infection en début de lactation (caractères de production également mesurés) et différentiels entre les individus porteurs de la mutation et ceux non porteurs. Pour ce faire, j'ai conduit une expérimentation basée sur un challenge expérimental chez des brebis portant les génotypes *SOCS2* extrêmes. J'ai ensuite analysé les nombreuses données générées (phénotypiques, cliniques, et transcriptomiques) à l'aide d'analyses statistiques variées telles que des modèles linéaires, des analyses d'expression des gènes, ou encore des analyses multivariées permettant l'intégration de l'ensemble des données.

Ce manuscrit débutera par une étude bibliographique ([Chapitre I.](#)) qui présentera dans un premier temps le contexte scientifique entourant la prise en compte de mutations dans les schémas de sélection en élevage. Un état des lieux sur les connaissances concernant le gène *SOCS2* et ses fonctions biologiques sera ensuite réalisé. Cette partie bibliographique se clôturera par un article de synthèse bibliographique que j'ai rédigé concernant les études génomiques sur les mammites ovines, et qui met en lumière les relations entre ce caractère de santé génétiquement complexe et les autres caractères d'intérêt pour la filière. Les démarches, résultats et discussion des différents objectifs cités ci-dessus seront ensuite déclinés. Ces chapitres seront présentés soit sous la forme d'articles scientifiques acceptés pour publication ([Chapitres II.](#) et [III.](#)), pouvant être accompagnés de résultats et discussions complémentaires, soit sous forme de parties structurées scientifiquement avec des résultats non encore publiés ([Chapitre IV.](#)). Ce manuscrit s'achèvera par une discussion générale sur les travaux ainsi que des perspectives pour la filière ([Chapitre V.](#)).

Chapitre I. Introduction bibliographique

I. Gestion de mutations dans les schémas de sélection

Nous débuterons cette partie par nous intéresser à l'origine et à la nature de la variabilité génétique qui nous permet de réaliser une sélection des individus en élevage, avec notamment la notion de mutation qui est au cœur de ce sujet de thèse. Puis, nous présenterons les principes généraux de l'amélioration génétique des espèces d'élevage. Enfin, nous nous concentrerons sur les gènes majeurs et mutations causales connus à ce jour chez les ruminants et à leur prise en compte dans les schémas de sélection, et nous terminerons par la description du cas particulier de la mutation à l'étude, la mutation R96C dans le gène *SOCS2* chez la brebis laitière.

I.1. Origine, nature et conséquences de la variabilité génétique

La sélection des animaux d'élevage est permise par l'existence d'une variabilité dans le génome des individus au sein d'une population, se traduisant par des différences phénotypiques. Nous allons dans cette partie présenter cette notion de variabilité génétique.

I.1.a. Définitions générales de génétique

L'information génétique des êtres vivants est contenue dans l'ADN (Acide DésoxyriboNucléique), présent dans le noyau des cellules eucaryotes. Une molécule d'ADN (ou polymère nucléotidique) est composée de deux brins antiparallèles liés au niveau de bases alléliques (ou nucléotides) (Figure 1). Ces nucléotides contiennent les bases azotées (A pour Adénine, C pour Cytosine, G pour Guanine, et T pour Thymine), liées par paires (A avec T, C avec G) au sein d'une structure en double hélice. On peut avoir accès à la connaissance de cette succession de nucléotides (ou séquence) grâce au séquençage.

Sur la séquence du génome humain, environ 98% des régions sont non codantes. Autrement dit, seulement 2% de l'ADN correspond à des gènes qui vont être transcrits en ARN (Acide RiboNucléique) messenger, ou ARNm, puis traduits en protéines (succession d'acides aminés) qui sont les principales composantes des structures des cellules et à la base de nombreuses hormones, enzymes et anticorps nécessaires au bon fonctionnement de l'organisme. Chez l'Homme par exemple, on compte aujourd'hui un peu plus de 20 000 gènes connus codant pour des protéines (Salzberg, 2018). Les régions non codantes restent encore très partiellement connues dans l'ensemble des espèces mais servent principalement à la régulation de l'expression des gènes.

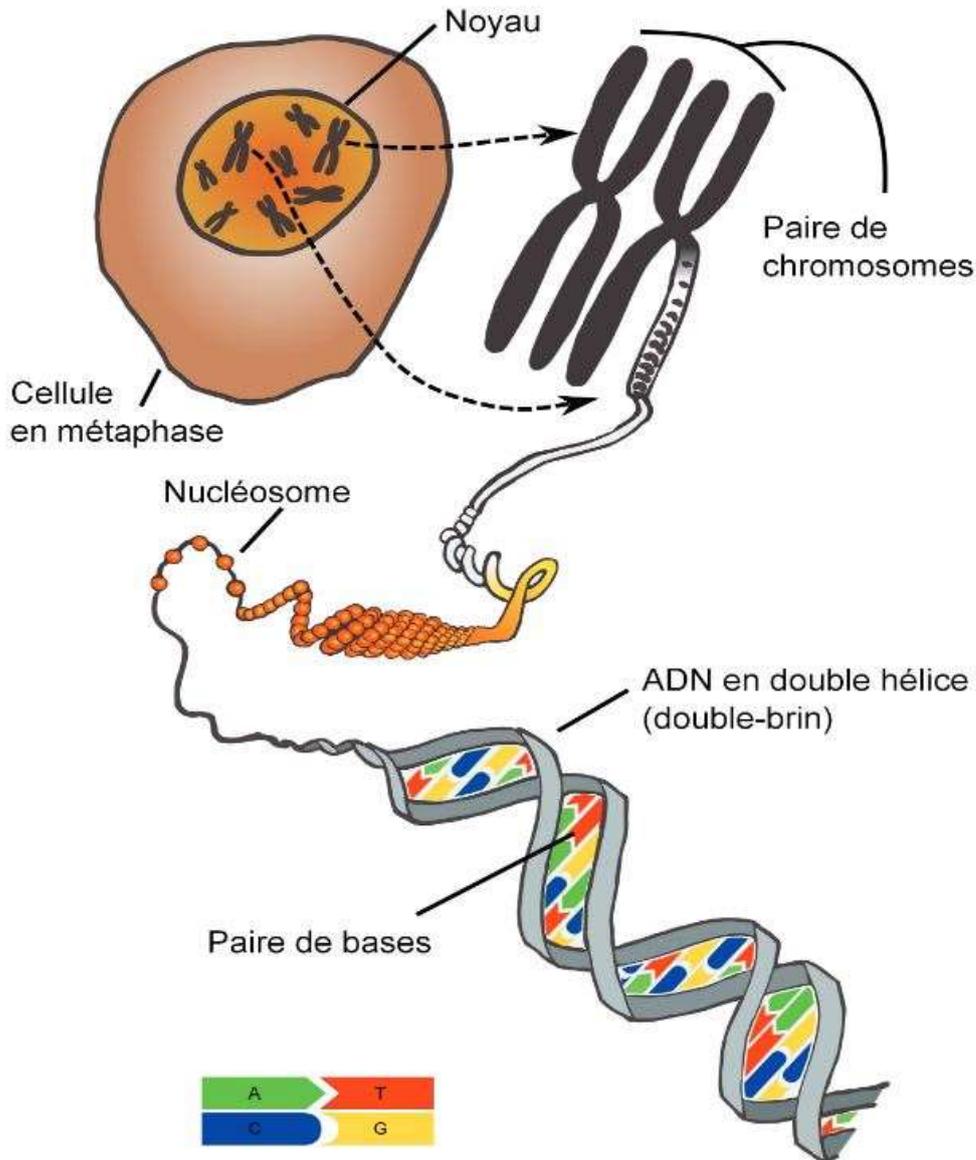


Figure 1 Illustration schématique de l'information génétique dans les cellules eucaryotes (Vignal, 2011)

Abréviations : A : Adénine ; ADN : Acide DésoxyriboNucléique ; C : Cytosine ; G : Guanine ; T : Thymine.

Chez les mammifères, l'ADN est stocké dans le chromosome (Figure 1) (structure compacte de protection) qui est présent en un seul exemplaire dans les cellules sexuelles (gamètes), et en deux exemplaires (origines paternelle et maternelle) dans toutes les autres cellules de l'organisme. Selon l'espèce, le nombre de chromosomes et leur longueur varient. Par exemple, l'Homme possède 46 chromosomes (2 x 22 autosomes + 2 chromosomes sexuels) dont la longueur totale est d'environ 3 milliards de paires de nucléotides (3 000 Mb *i.e.* Mégabase). Ainsi, tout individu possède ses nucléotides en deux exemplaires (provenant de chaque chromosome), chaque exemplaire étant appelé allèle ou base allélique. Cette paire d'allèles est le génotype de l'individu à une position du génome, ou *locus*. Le génotypage est l'accès à la connaissance des génotypes d'un individu dont les *loci* sont connus. Un génotype peut être

qualifié d'homozygote, lorsque ses deux allèles sont identiques, ou hétérozygote, lorsque ses deux allèles sont différents. Tous les *loci* du génome ne sont pas polymorphes (*i.e.* qui possèdent plusieurs bases alléliques) : il existe des portions du génome où tous les individus d'une population sont homozygotes. Lorsqu'un *locus* est polymorphe, on parle alors de SNP (*Single Nucleotide Polymorphism* - polymorphisme nucléotidique), marqueur moléculaire très utilisé en génétique quantitative pour le génotypage à grande échelle.

I.1.b. Les différentes sources de variabilité génétique

La variabilité génétique présente au sein d'une population résulte de la modification naturelle du génome des individus d'une génération à une autre, c'est-à-dire que le génome du descendant n'est pas exactement une combinaison du génome de ses parents : il y a un phénomène de mutation. Les mutations qui affectent la descendance, principalement les mutations dites germinales, c'est-à-dire présentes dans une cellule sexuelle (gamète) d'un individu (il existe aussi plus rarement des mutations post-zygotiques apparaissant après fécondation, pouvant affecter tout ou partie des cellules du nouvel organisme), sont à distinguer des mutations somatiques qui affectent toutes les autres cellules et qui sont à l'origine de nombreux cancers. Les mutations peuvent être induites par des agents extérieurs mutagènes (rayons ultraviolets ou électromagnétiques par exemple), ou spontanées (apparition aléatoire dans une cellule), et se divisent en trois classes :

- 1) Mutations génomiques : polyploïdisation, c'est-à-dire apparition de plus de deux chromosomes par paire (entraînant souvent un avortement naturel),
- 2) Mutations chromosomiques, aussi appelées variants de structure : duplication, translocation, inversion, délétion, ou recombinaison d'une partie de chromosome,
- 3) Mutations géniques, ou ponctuelles : modification dans la succession des bases azotées (petit nombre de bases touchées) dans une région codante.

Nous allons nous attarder sur les mutations ponctuelles, au cœur de ce sujet de thèse, qui sont principalement dues à des erreurs de répllication de l'ADN, soit non détectées par les systèmes de réparation, soit mal réparées. On distingue deux types de mutations ponctuelles : les mutations décalantes et les mutations par substitution (Figure 2).

SÉQUENCE DE DÉPART						
ADN	codons	TCG	TTT	GCA	CAC	
(dans transcrit)						
ARNm	codons	AGC	AAA	CGU	GUG	
	aa	Séine	Lysine	Arginine	Valine	
Protéine	abr. au 3 lettres	Ser	Lys	Arg	Val	
	abr. au 1 lettre	S	K	R	V	

MUTATIONS DÉCALANTES						
Insertion						
ADN	codons	AVANT	TCG	TTT	GCA	CAC
(dans transcrit)		APRÈS	TCG	TAT	TGC	ACA
ARNm	codons		AGC	AAA	ACG	UGU
	aa	Séine	Isolectine	Thréonine	Cystéine	
Protéine	abr. au 3 lettres	Ser	Ile	Thr	Cys	
	abr. au 1 lettre	S	I	T	C	
Délétion						
ADN	codons	AVANT	TCG	TTT	GCA	CAC
(dans transcrit)		APRÈS	TCG	TTG	CAC	ACC
ARNm	codons		AGC	AAC	GAG	UGG
	aa	Séine	Asparagine	Acide glutamique	Tryptophane	
Protéine	abr. au 3 lettres	Ser	Asn	Glu	Trp	
	abr. au 1 lettre	S	N	E	W	

MUTATIONS PAR SUBSTITUTION						
Silencieuse						
ADN	codons	AVANT	TCG	TTT	GCA	CAC
(dans transcrit)		APRÈS	TCG	TTG	GCA	CAC
ARNm	codons		AGC	AAC	CGU	GUG
	aa	Séine	Lysine	Arginine	Valine	
Protéine	abr. au 3 lettres	Ser	Lys	Arg	Val	
	abr. au 1 lettre	S	K	R	V	
Faux-sens						
ADN	codons	AVANT	TCG	TTT	GCA	CAC
(dans transcrit)		APRÈS	TCG	TTT	ACA	CAC
ARNm	codons		AGC	AAA	UGU	GUG
	aa	Séine	Lysine	Cystéine	Valine	
Protéine	abr. au 3 lettres	Ser	Lys	Cys	Val	
	abr. au 1 lettre	S	K	C	V	
Non-sens						
ADN	codons	AVANT	TCG	TTT	GCA	CAC
(dans transcrit)		APRÈS	TCG	ATT	GCA	CAC
ARNm	codons		AGC	UAA	CGU	GUG
	aa	Séine	STOP	-	-	-
Protéine	abr. au 3 lettres	Ser	-	-	-	-
	abr. au 1 lettre	S	-	-	-	-

Figure 2 Les différents types de mutations ponctuelles et leur conséquence sur la séquence protéique, quelques exemples concrets

Abréviations : A : Adénine ; aa : Acide Aminé ; abr. : abréviation ; ADN : Acide DésoxyriboNucléique ; ARNm : Acide RiboNucléique messager ; C : Cytosine ; G : Guanine ; T : Thymine ; U : Uracile.

Les mutations décalantes (insertion ou délétion) sont des mutations qui décalent le cadre de lecture. Une insertion (délétion) est l'ajout (la suppression) d'une base nucléotidique dans la séquence ADN. Les mutations par substitution sont quant à elles au nombre de trois :

- 1) Silencieuse : modification d'un nucléotide sans modification de la protéine correspondante grâce à la redondance du codage des acides aminés,
- 2) Faux-sens : modification d'un nucléotide modifiant l'acide aminé,
- 3) Non-sens : modification d'un nucléotide aboutissant à un codon STOP (traduction stoppée et protéine tronquée, donc modifiée).

Les mutations ponctuelles qui entraînent la modification d'un acide aminé peuvent être problématiques car la séquence protéique va être modifiée et peut entraîner des pertes de fonctionnalité de la protéine.

Les mutations ponctuelles sont des événements rares à l'échelle du nucléotide. Chez l'Homme, le taux de mutations ponctuelles par substitution a été estimé à $3,0 \times 10^{-8}$ mutations/nucléotide/génération (Xue *et al.*, 2009). En revanche, à l'échelle du génome (c'est-à-dire 3 000 Mb), en appliquant ce taux de mutation, on obtient un nombre approximatif de 90 mutations par gamète, soit environ 180 nouvelles mutations par nouvel individu. A chaque

Page | 25

génération, un nombre conséquent de nouvelles mutations est donc produit, faisant ainsi apparaître de nouveaux génotypes qui peuvent avoir des conséquences favorables ou défavorables sur un caractère donné. Une mutation dans une région non codante peut également avoir un effet direct sur la détermination des caractères dans le cas où le *locus* de la mutation est en déséquilibre de liaison avec un gène codant (association non aléatoire entre *loci*, c'est-à-dire qu'ils vont être transmis en même temps à la descendance).

1.1.c. Architecture génétique des caractères

En génétique quantitative, l'investigation du déterminisme génétique des caractères est un sujet majeur de recherche, et s'est accélérée depuis l'arrivée des nouveaux outils de génotypage à grand échelle que sont les puces à SNP à partir des années 1990. Ces outils permettent en effet de connaître simultanément le génotype de nombreux *loci* à la fois, de l'ordre de quelques milliers de SNP pour les puces basse densité à quelques centaines de milliers pour les puces haute densité.

Le modèle génétique classique décompose la performance d'un individu (*i.e.* la mesure du caractère en question) en un effet dû à la génétique et un effet dû au milieu environnemental (en supposant qu'il n'y ait pas d'interaction entre ces deux effets) (Fisher, 1918). Ce modèle, appelé modèle polygénique infinitésimal, part du principe que les effets génétiques sont dus à une multitude de positions du génome à effets faibles, c'est-à-dire que les caractères quantitatifs sont contrôlés par un grand nombre de gènes ayant chacun un effet faible. Plus la part de la variabilité phénotypique d'un caractère est expliquée par l'effet génétique, plus le caractère est dit héritable, et plus la descendance aura une performance pour ce caractère similaire à celle de ses parents. L'héritabilité d'un caractère correspond à la part de variance expliquée par cet effet génétique par rapport à la variance totale du caractère (valeur comprise entre 0 et 1). Deux caractères peuvent parfois être liés génétiquement, on parle alors de corrélation génétique. Cette corrélation génétique peut s'expliquer par : (i) des mêmes régions du génome ou gènes impliqués dans la détermination de plusieurs caractères (on parle alors de pléiotropie), ou bien, (ii) des régions du génome ou gènes différents en déséquilibre de liaison et impliqués dans la détermination des différents caractères.

La recherche du déterminisme génétique des caractères consiste donc à trouver les régions du génome à l'origine d'un caractère, c'est-à-dire les QTL (*Quantitative Trait Locus* - *locus* de caractère quantitatif), et, si possible, les gènes et mutations causaux sous-jacents. Cette détection de QTL se fait classiquement à l'aide de tests statistiques d'association entre des performances et les marqueurs moléculaires SNP à l'aide de modèles de régression linéaires.

Par ailleurs, on tient compte dans ces modèles de la structure de la population et de l'effet du fond génétique à l'aide d'une matrice de parenté (génomique ou basée sur la connaissance du pedigree). Les SNP sont testés un à un afin de détecter quels polymorphismes sont à l'origine de la variabilité du phénotype étudié. Lorsque les SNP disponibles sont répartis sur l'ensemble du génome, on parle de GWAS (*Genome-Wide Association Studies* - études d'associations pangénomiques). Dans la plupart de ces études, les SNP causaux ne comptent pas parmi les marqueurs disponibles et l'on teste des SNP dits neutres, c'est-à-dire qu'ils ne sont pas en association directe avec le phénotype. En revanche, grâce au déséquilibre de liaison, ces SNP neutres peuvent être liés aux SNP causaux. Pour cette raison, ils peuvent ressortir associés au caractère dans les GWAS et permettre la localisation précise de la région (avec un intervalle de confiance) du gène ou du SNP causal. Les données de séquençage sont prometteuses pour la détection directe des mutations causales dans les études d'association mais restent encore aujourd'hui d'un coût élevé.

Ces approches GWAS avec l'utilisation de puces à SNP ont mis en évidence la complexité des caractères quantitatifs déterminés par de multiples QTL à effets plus ou moins faibles, validant ainsi avec brio la théorie du modèle polygénique infinitésimal développée un siècle auparavant (Visscher & Goddard, 2019). Par exemple, le phénotype de taille chez l'homme (héritabilité ~ 0,8) a été découvert associé, grâce à un jeu de données comprenant plus de 700 000 individus européens, à plus de 3000 *loci* qui n'expliquent pourtant que 24,6% de la variance génétique (Yengo *et al.*, 2018). Des QTL dits « à effets forts » sont des QTL au sein desquels on peut découvrir des gènes majeurs c'est-à-dire des gènes qui expliquent une plus grande part de la variance génétique. L'existence de ces gènes majeurs ne remet pas en cause la composante polygénique (autres *loci* déterminants à effet plus faibles) mais est à prendre en considération et notamment dans les espèces d'élevage où la fixation de certains allèles de gènes majeurs peut conduire à une perte locale de diversité.

I.2. Principes généraux de l'amélioration génétique des espèces d'élevage

I.2.a. Principe général des schémas de sélection

La sélection artificielle, ou dirigée par l'Homme, des espèces d'élevage est effectuée depuis le début de la domestication des animaux de rente au cours du Néolithique (Zeder, 2008). En effet, les populations domestiques ont génétiquement évolué différemment des populations restées à l'état sauvage en raison d'une part de l'adaptation à leur nouvel environnement domestique et d'autre part selon les besoins des éleveurs (animaux conservés et mis à la reproduction étant les plus productifs, résistants, et faciles à manipuler). De cette manière, des phénotypes favorables

ont vu leur fréquence augmenter dans certaines régions géographiques selon les différentes pratiques d'élevage et conditions environnementales, aboutissant à la différenciation de plusieurs populations au sein d'une même espèce, appelées ultérieurement races.

Ce n'est qu'à partir du XIX^{ème} siècle en France que des organisations d'éleveurs se sont créées pour permettre un gain génétique plus grand au sein de leurs troupeaux et de la population globale, grâce notamment à la mise en place de livres généalogiques par race chez les bovins à partir des années 1850 (1855 pour la race Durham, 1922 pour la race Prim'Holstein anciennement appelée Hollandaise, etc.). Ont suivi les espèces caprines et ovines (1930 pour la race caprine Alpine, 1942 pour la race ovine Lacaune). Grâce aux enregistrements généalogiques, ces livres permettaient également aux éleveurs de certifier de la « pureté » de leurs animaux qui devaient se conformer au standard de la race (critères phénotypiques) et être identifiés. Ces trois espèces sont aujourd'hui sélectionnées au sein de nombreux Organismes de Sélection (OS) qui mettent en œuvre des programmes d'amélioration génétique, ou schémas de sélection, pour une race, l'ensemble de ces OS étant coordonné par l'Institut de l'Élevage (Idele), créé en 1991 à partir de la fusion de deux Instituts, l'ITEB (Institut Technique de l'Élevage Bovin) créé en 1962, et l'ITOVIC (Institut Technique de l'élevage OVIN et Caprin) créé en 1968. De la même manière, des institutions se sont créées pour l'organisation des autres filières avec le SYSAAF (Syndicat des Sélectionneurs Avicoles et Aquacoles) pour les filières aquacoles et avicoles, créé en 1952 sous le nom de SNAA (Syndicat National des Aviculteurs Agréés), et l'IFIP (Institut du porc) créé en 2006 par fusion du CTSCCV (Centre Technique de la Salaison, de la Charcuterie et des Conserves de Viandes), créé en 1954, et de l'ITP (Institut Technique du Porc), lui-même créé en 1961.

Le principe de base d'un schéma de sélection (Figure 3), quelle que soit la filière, repose sur l'élevage d'un ensemble d'individus phénotypés et évalués quantitativement sur des caractères d'intérêt. Parmi ces individus, la semence de ceux obtenant les meilleures performances répondant à des objectifs de sélection bien définis par la filière va être disséminée à travers l'ensemble de la population. Au sein de ces ensembles d'individus, les accouplements sont raisonnés de manière à limiter les pertes de diversité génétique dans la population globale (*i.e.* augmentation de la consanguinité). On parle souvent d'organisation pyramidale des schémas de sélection en raison du nombre restreint d'animaux à l'origine des descendants de la majorité de la population. Dans les filières avicoles et porcines, le haut de la pyramide correspond à des individus de lignées pures ayant des phénotypes connus et définis. Les sélections des meilleurs reproducteurs sont réalisées dans ces lignées pures qui sont croisées pour obtenir une nouvelle population possédant les caractéristiques à la fois paternelles et maternelles. Le niveau des

« Multiplicateurs » est ensuite la production des individus parentaux qui, après croisement, seront à l'origine des individus élevés et commercialisés. Dans les filières bovines, ovines et caprines, les schémas de sélection s'effectuent par race avec un noyau de sélection qui propose directement aux éleveurs des mâles reproducteurs améliorateurs (semence et/ou monte naturelle), sélectionnés sur leurs performances ou sur ascendance.

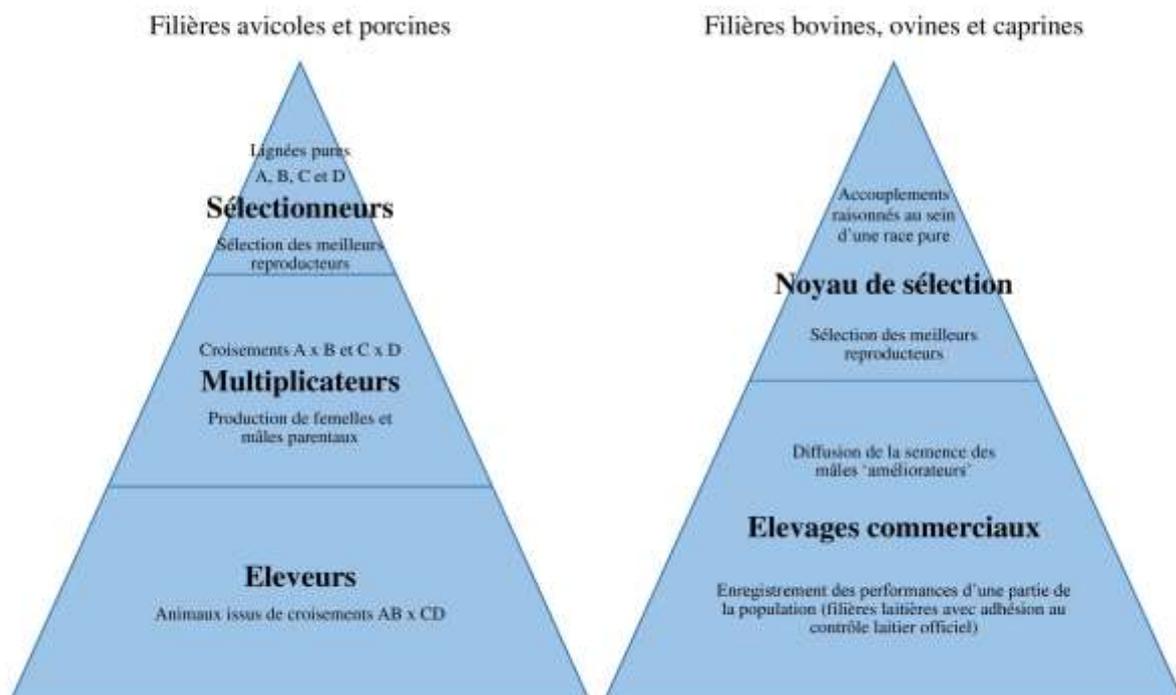


Figure 3 Représentation schématique de l'organisation des programmes de sélection français dans les différentes filières d'animaux de rente

I.2.b. L'objectif de sélection et les index

Les éleveurs peuvent eux-mêmes effectuer, individuellement, une sélection des meilleurs animaux sur leurs propres performances d'intérêt au sein du troupeau, ce qui peut permettre un gain génétique peu maîtrisé au cours des générations. C'est ce qu'on appelle la sélection massale. Afin de mieux maîtriser ce gain génétique (rapidité, gestion de la variabilité génétique, etc.), l'ensemble des acteurs de la filière (éleveurs, transformateurs, consommateurs, etc.) déterminent collectivement un objectif de sélection. Cet objectif de sélection correspond aux caractères à améliorer au sein d'une population qui répondent donc à des enjeux sociétaux (se nourrir, se couvrir, enjeu environnemental, bien-être animal), liés aux éleveurs (production, santé, conformation, comportement, enjeu économique), ainsi qu'aux industriels (enjeu économique).

Ces caractères sont choisis parmi les caractères héréditaires. En effet, la réponse à la sélection (Δ), dépend de l'hérédité du caractère ou plus précisément de la variance génétique (σ_g), de l'intensité de la sélection (i), de la précision des évaluations génétiques (ρ), et de l'intervalle de génération (T), selon la formule suivante :
$$\Delta = \frac{\sigma_g \times i \times \rho}{T}$$
. L'intensité de sélection correspond au taux d'individus retenus dans le noyau de sélection (haut de la pyramide) par rapport au nombre de candidats possibles. Une intensité de sélection forte est favorable au progrès génétique mais représente un risque sur la variabilité génétique globale de la population. La précision des évaluations génétiques correspond à la corrélation entre les valeurs génétiques dites vraies (calculées à partir des performances de la descendance) et celles prédites par les modèles d'évaluation que nous allons aborder dans la section suivante. L'intervalle de génération est défini comme l'âge moyen des reproducteurs à la naissance de leurs descendants et est fixe au sein d'un schéma de sélection.

Chez les animaux d'élevage, les principaux caractères en sélection sont des caractères de production (viande, lait, œufs) avec des notions de quantité, de qualité, et d'efficacité (produire plus, de meilleure qualité, et plus rapidement). A ces caractères s'ajoutent des caractères de reproduction (aptitudes maternelles, qualité de semence paternelle) et des caractères fonctionnels (qui ne rapportent pas directement mais améliorent la rentabilité en limitant les coûts de production : santé, morphologie mammaire, longévité, etc.). De nombreuses études sont également en cours pour l'intégration de caractères issus des thématiques sociétales de réduction des rejets de gaz à effets de serre dans l'environnement (Hayes *et al.*, 2016 ; Quinton *et al.*, 2018 ; Barwick *et al.*, 2019) et d'amélioration du bien-être animal (Nielsen *et al.*, 2017 ; Zuliani *et al.*, 2018).

La théorie des indices de sélection (Hazel, 1943) est à l'origine des valeurs appelées index synthétiques, mises à disposition des éleveurs par les OS pour choisir les meilleurs reproducteurs. Cette théorie repose sur un objectif de sélection H multi-caractères qui est une combinaison linéaire de valeurs génétiques (G) pour les n caractères d'intérêt définis par la filière, pondérées par leur importance économique : $H = a_1 G_1 + a_2 G_2 + \dots + a_n G_n$, avec a_i la pondération économique du $i^{\text{ième}}$ caractère. En pratique, cet objectif de sélection H n'est pas calculable car les valeurs génétiques G_i ne sont pas connues, mais estimable à l'aide de phénotypes indicateurs observables dans la population. Cette estimation de H est appelée index (I), qui est alors une combinaison linéaire de m phénotypes observables (X_j) pré-

corrigés pour les effets du milieu et centrés : $I = b_1X_1 + b_2X_2 + \dots + b_mX_m$, avec b_j la pondération du $j^{\text{ième}}$ phénotype. Le but étant d'approcher au mieux H , les pondérations b_j sont choisies de manière à maximiser la corrélation entre H et I . Le vecteur des pondérations (b) qui maximise la corrélation entre H et I est : $b = P^{-1}Qa$, où P désigne la matrice de variance-covariance des m performances, Q , la matrice de variance-covariance entre les n valeurs génétiques et les m performances, et a , le vecteur des pondérations économiques des n caractères. Il est à noter qu'un caractère de l'objectif de sélection peut être expliqué par plusieurs phénotypes observables dans la population et n n'est donc pas nécessairement égal à m . Ce type d'index synthétique est plus efficace que la sélection massale et permet un gain génétique optimisé pour chacun des caractères de l'objectif de sélection dans la population globale. La sélection est donc un subtil compromis entre le poids économique des caractères d'intérêt, leur héritabilité et les corrélations génétiques entre ces caractères.

En race ovine Lacaune lait en France par exemple, est proposé un index synthétique appelé ISOL (Index Synthétique Ovin Lait) (Barillet *et al.*, 2016) dans lequel sont pondérés à hauteur de 50% de l'index les caractères laitiers (quantité de lait, de matières et taux protéiques et butyreux), et l'autre moitié est une combinaison des caractères fonctionnels (CCS et morphologie mammaire). L'ISOL a permis un gain de +4,1 L de lait, +0,17 g/L de taux butyreux et +0,16 g/L de taux protéique par an, entre 2008 et 2015 en moyenne en race Lacaune lait (Barillet *et al.*, 2016).

Il existe des limites à l'efficacité de cette sélection notamment en raison de l'antagonisme entre les différents caractères en sélection (corrélations génétiques négatives). Classiquement sont opposés les caractères de production avec les caractères fonctionnels (santé, longévité, etc.) (Pritchard *et al.*, 2013) et également de reproduction (Mäntysaari & Van Vleck, 1989). Cette opposition s'est illustrée dans un premier temps dans de nombreuses races et espèces par une amélioration du premier caractère d'intérêt sélectionné dans les programmes d'amélioration génétique et une détérioration des autres caractères, comme c'est le cas par exemple en Suède en bovin lait où la production laitière a été fortement sélectionnée à partir des années 60 et où le caractère de santé mammaire n'a été pris en compte dans le programme de sélection qu'à partir de 1985 suite à sa dégradation (Lindhé & Philipsson, 1998). Cette corrélation génétique peut s'expliquer par : (i) des mêmes régions du génome ou gènes impliqués dans la détermination de plusieurs caractères antagonistes (on parle alors de pléiotropie), ou bien (ii) des gènes différents en déséquilibre de liaison impliqués dans la détermination de caractères antagonistes. Dans le second cas (ii), il est possible de pallier ce problème en cassant cette

liaison, c'est-à-dire en découplant les deux *loci* par exemple par croisement successif ou encore par édition du génome. Cela permettrait de pouvoir sélectionner séparément les deux caractères initialement liés de manière antagoniste. Le premier cas est plus problématique car le découplage n'est pas possible, la question du compromis entre les caractères est alors en jeu et il faut alors pouvoir évaluer les effets du gène sur ces différents caractères.

1.2.c. L'évaluation génétique/génomique

Nous venons de voir que les index synthétiques proposés par les OS aux éleveurs sont des combinaisons de valeurs génétiques estimées d'un phénotype observable pour un individu. Le but de l'évaluation génétique est donc d'estimer une valeur génétique quantitative pour un caractère donné pour un individu, valeur qui vise à corriger l'observation (phénotype) pour les effets environnementaux (milieu de vie de l'animal, état physiologique, facteurs techniques de prise des mesures, etc.). Grâce à cette estimation, des individus élevés avec des pratiques d'élevage différentes, et/ou dans une période temporelle différente, et/ou dont la performance est mesurée avec des outils ou des observateurs différents peuvent être comparés. Cette estimation se fait grâce à la connaissance de valeurs phénotypiques enregistrées pour un grand nombre d'individus plus ou moins apparentés.

Les valeurs génétiques sont classiquement estimées à l'aide de modèles statistiques basés sur le modèle infinitésimal (Fisher, 1918) qui décomposent la performance (phénotype) en effets dus à l'environnement, et en effets dus au patrimoine génétique de l'individu (on parle de valeur génétique additive qui correspond à la somme de l'ensemble des valeurs individuelles des positions du génome qui détermineraient le caractère). Le modèle plus utilisé est un modèle animal appelé BLUP (*Best Linear Unbiased Predictor* - meilleure prédiction linéaire non biaisée) (Henderson, 1975) qui est basé sur un modèle linéaire mixte et qui permet une évaluation conjointe de ces effets de milieu et génétiques. Ce modèle part donc du principe que les effets génétiques sont dus à une multitude de positions du génome à effets faibles et ne prend pas en compte la présence de gènes majeurs pour le déterminisme génétique des caractères.

Initialement, ces valeurs génétiques étaient estimées à partir des performances enregistrées, et des relations de parenté entre les individus, calculées à partir de la généalogie (pedigree) connue, sans a priori sur les données génétiques propres des individus. Le développement des techniques de séquençage des génomes et de génotypage, notamment avec l'arrivée des puces à SNP a stimulé l'émergence de modèles d'évaluation permettant d'inclure l'information génomique qui est alors utilisée pour mieux calculer les relations de parenté entre individus (Lande & Thompson, 1990 ; Meuwissen *et al.*, 2001). Cet apport génomique a augmenté les

précisions des valeurs génétiques prédites et ont permis une sélection plus efficace au sein des espèces d'élevage qui ont pu mettre en place ces dispositifs (Boichard *et al.*, 2016). En revanche, ces modèles ne considèrent pas de façon optimale la présence de QTL à effets forts ou de gènes majeurs pour un caractère dans le génome car ils considèrent que les effets des SNP sont tous issus de la même distribution normale (Goddard *et al.*, 2016).

Dans un premier temps, une méthode simple pour prendre en compte cette information consistait à inclure le génotypage du QTL comme effet fixe dans le modèle BLUP (Kennedy *et al.*, 1992). Mais cette méthode est très contraignante car elle nécessite le génotypage de tous les individus avec performance, et les mutations causales font rarement partie des SNP présents dans une puce. Une méthode plus récente appelée *Gene Content* a été développée pour prendre en compte le génotypage du gène majeur/QTL en tant que caractère dans une approche multi-caractère (Legarra & Vitezica, 2015). Cette méthode est adaptée pour le cas des populations partiellement génotypées. Une autre approche consiste à estimer au préalable les effets des marqueurs à partir des solutions du BLUP et à entrer en paramètre d'entrée cette pondération afin de donner plus de poids aux marqueurs associés au caractère évalué (Wang *et al.*, 2012 ; Zhang *et al.*, 2016). En outre, des approches bayésiennes telles que la méthode appelée Bayes R (Erbe *et al.*, 2012) permettent également de prendre en compte le fait que tous les SNP ne sont pas associés à des QTL en donnant des distributions *a priori* différentes en paramètre d'entrée. Cependant, en petits ruminants, il a été montré que les méthodes bayésiennes ne donnaient pas de meilleures précisions pour les valeurs génétiques estimées (Duchemin *et al.*, 2012) et l'approche usuellement utilisée est le BLUP.

Tous les modèles développés pour prendre en compte ces gènes à effet majeur dans les approches BLUP ne sont pas encore implémentés en routine dans les schémas de sélection à ce jour. Pour le moment, les mesures mises en place pour sélectionner des gènes majeurs restent majoritairement des sélections d'individus améliorateurs qui possèdent la forme du gène majeur la plus favorable (connaissance du gène directement), en parallèle de leur valeur génétique estimée.

I.3. Inventaire et prise en compte dans les schémas de sélection des gènes majeurs et mutations causales en ruminants

Nous allons à présent nous intéresser à la prise en compte dans les schémas de sélection des gènes majeurs pour des caractères d'intérêt chez les ruminants.

1.3.a. Synthèse des principales mutations causales de gènes majeurs dans les filières laitières et viande chez les ruminants

Les caractères d'intérêt en élevage sont complexes et déterminés à la fois par des facteurs environnementaux et par des facteurs génétiques. Le modèle infinitésimal mentionné précédemment suppose que les caractères sont déterminés par un grand nombre de *loci* du génome. Ceci peut s'expliquer biologiquement par la complexité des différentes fonctions de l'organisme animal qui nécessitent l'interaction et l'action conjointe de nombreuses molécules codées par différents gènes, expliquant chacun une part faible de la variance génétique. Au sein de ces mécanismes biologiques, certaines molécules peuvent parfois jouer des rôles plus importants que les autres pour la caractérisation du phénotype d'intérêt. Une modification du fonctionnement de telles molécules, dues par exemple à une mutation causale dans son gène codant, peut donc entraîner des variations (amélioration ou défaut) du phénotype déterminé.

Les gènes majeurs qui expliquent donc une grande part de la variance génétique pour la détermination d'un caractère sont assez peu fréquents pour les caractères d'intérêt en élevage en raison de leur complexité. En effet, les fonctions de production, de reproduction, ou encore de résistance aux maladies sont complexes, en revanche, plus un phénotype est précis, plus grand est le risque de trouver un gène majeur en raison du fait que l'on réduit l'amplitude de fonctions biologiques à investiguer. La détection de QTL a permis depuis les années 1990 de cartographier le déterminisme génétique des phénotypes d'intérêt en élevage (Goddard *et al.*, 2016) et notamment de localiser des gènes majeurs contenant des mutations causales. Une liste non exhaustive de mutations causales dans des gènes majeurs de caractères d'intérêt des filières laitières et allaitantes chez les ruminants est présentée dans le Tableau 1.

Malgré les fortes héritabilités des caractères de production, peu de gènes majeurs sont décrits dans la littérature (Tableau 1). En effet, pour les filières laitières, les gènes majeurs connus déterminent des phénotypes bien précis comme par exemple le taux d'une molécule dans le lait, comme c'est le cas pour les nombreux variants connus des caséines (Caroli *et al.*, 2009 ; Selvaggi *et al.*, 2014a ; Selvaggi *et al.*, 2014b), protéines d'importance économique forte en raison de leur implication dans le procédé de transformation du lait en fromage. Pour les filières allaitantes, un gène majeur connu, codant pour la myostatine, est associé au phénotype « culard » qui correspond à une hypertrophie musculaire de l'arrière-train. Des mutations causales sont connues dans ce gène en bovin (Grobet *et al.*, 1997 ; Kambadur *et al.*, 1997) et en ovin (Clop *et al.*, 2006).

Tableau 1 Principales mutations causales dans des gènes majeurs de caractères d'intérêt des filières laitières et allaitantes chez les ruminants

<i>Gène majeur</i>	<i>Phénotype associé</i>	<i>Chromosome : mutation(s) causales connues</i>	<i>Population(s) de découverte</i>	<i>Références</i>
PRODUCTION				
<i>ABCG2</i>	Lait et taux	BTA6 : p.Y581S	Holstein	(Cohen-Zinder <i>et al.</i> , 2005)
<i>BCO2</i>	Couleur du tissu adipeux	OAR15 : c.196C>T	Norwegian White	(Våge & Boman, 2010)
<i>CSN1S1</i>	Taux protéique du lait	BTA6/CHI6/OAR6 : ≥ 9/17/8 variants connus	Nombreuses races	(Caroli <i>et al.</i> , 2009)
<i>CSN1S2</i>		≥ 4/8/7 variants connus	Nombreuses races	(Marletta <i>et al.</i> , 2007 ; Selvaggi <i>et al.</i> , 2014b)
<i>CSN2</i>		≥ 9/8/5 variants connus	Nombreuses races	(Selvaggi <i>et al.</i> , 2014a)
<i>CSN3</i>		≥ 11/16/1 variants connus		
<i>DGAT1</i>	Taux butyreux du lait	BTA14 : p.K232A CHI14 : p.R251L ; p.R396W	Nombreuses races	(Grisart <i>et al.</i> , 2002 ; Winter <i>et al.</i> , 2002) (Martin <i>et al.</i> , 2017)
<i>GPIHBP1</i>		BTA14 : g.2553998G>A	Holstein	(Yang <i>et al.</i> , 2017)
<i>MSTN/GDF8</i>	Hypertrophie musculaire	BTA2 : c.821del11 c.941G>A OAR2 : g+6723G-A	Blanc Bleu Belge Piémontaise Texel	(Grobet <i>et al.</i> , 1997 ; Kambadur <i>et al.</i> , 1997) (Clop <i>et al.</i> , 2006)

REPRODUCTION

<i>B4GALNT2</i>	Taux d'ovulation / Prolificité	OAR11 : g.36938224T>A ; g.36938224T>A	Lacaune	(Drouilhet <i>et al.</i> , 2013)
<i>BMP15</i>		OARX : c.718C>T ; c.1100G>T c.871C>T ; c.896T>A c.962G>A c.525_541del c.1009A>C ; c.950C>T c.[301G>T ; 302_304del ; 310ins]	Belclare/Cambridge ; Belclare Hanna ; Inverdale Lacaune Rasa Aragonesa Grivette ; Olkuska Tunisian Barbarine	(Hanrahan <i>et al.</i> , 2004) (Galloway <i>et al.</i> , 2000) (Bodin <i>et al.</i> , 2007) (Martinez-Royo <i>et al.</i> , 2008) (Demars <i>et al.</i> , 2013) (Lassoued <i>et al.</i> , 2017)
<i>BMPR1B</i>		OAR6 : c.746A>G	Booroola Mérino	(Mulsant <i>et al.</i> , 2001 ; Souza <i>et al.</i> , 2001 ; Wilson <i>et al.</i> , 2001)
<i>GDF9</i>		OAR5 : c.1184C>T c.1279A>C c.1034T>G c.1111G>A c.943C>T	Belclare/Cambridge Icelandic Thoka Santa Inês Norwegian White ; Finn sheep Ile de France	(Hanrahan <i>et al.</i> , 2004) (Nicol <i>et al.</i> , 2009) (Silva <i>et al.</i> , 2011) (Våge <i>et al.</i> , 2013 ; Mullen & Hanrahan, 2014) (Souza <i>et al.</i> , 2014)

SANTÉ

<i>BoLA-DRB3</i>	Résistance aux infections	BTA23 : p.V53E ; p.V53L ; p.D57H ; p.D57N ; p.F60Y ; p.R84G	Nombreuses races	(Rastislav & Mangesh, 2012)
<i>ITGB2</i>	DAL	BTA1 : p.D128G	Holstein	(Shuster <i>et al.</i> , 1992)
<i>FEZL/FEZF2</i>	Mammites	BTA22 : p.G105(12_13)	Holstein	(Sugimoto <i>et al.</i> , 2006)
<i>PRNP</i>	Tremblante	OAR13 : p.A136V ; p.R154H	Nombreuses races	(Laplanche <i>et al.</i> , 1993)

		p.R171Q		(Goldmann <i>et al.</i> , 1990)
<i>SOCS2</i>	Inflammation mammaire	OAR3 : p.R96C	Lacaune	(Rupp <i>et al.</i> , 2015)
<i>TMEM154</i>	Visna-maëdi	OAR4 : p.L14H ; p.T25I ; p.D33N ; p.E35K ; p.T44M ; p.N70I	Nombreuses races	(Heaton <i>et al.</i> , 2012)

TARE GÉNÉTIQUE

<i>ABCA12</i>	Ichtyose	BTA2 : p. H1935R	Chianina	(Charlier <i>et al.</i> , 2008)
<i>APAF1</i>	Mortalité embryonnaire	BTA5 : p.Q579X	Holstein	(Adams <i>et al.</i> , 2016)
<i>APOB</i>	Déficit en cholestérol	BTA11 : p.Gly135FalfsX10	Holstein	(Kipp <i>et al.</i> , 2016)
<i>ATP2A1</i>	Dystonie	BTA25 : p.R559C	Blanc Bleu Belge	(Charlier <i>et al.</i> , 2008)
<i>CAD</i>	Mortalité embryonnaire	BTA11 : g.72399397T>C	Normande	(Mesbah-Uddin <i>et al.</i> , 2019)
<i>CWC15</i>		BTA15 : g.15707169C>T	Jersey	(Sonstegard <i>et al.</i> , 2013)
<i>FANCI</i>	Brachyspina	BTA21 : g.21184870_21188198del	Holstein	(Charlier <i>et al.</i> , 2012)
<i>GART</i>	Mortalité embryonnaire	BTA1 : p.N290T	Holstein	(Fritz <i>et al.</i> , 2013)
<i>ITGB4</i>	Épidermolyse bulleuse	OAR11 : c.4412_4415del	Churra	(Suárez-Vega <i>et al.</i> , 2015)
<i>MITF</i>	Microphthalmie	BTA22 : p.R211del	Holstein	(Bourneuf <i>et al.</i> , 2017)
<i>MRC2</i>	Queue tordue	BTA19 : c.2904_2905del c.1906T>C	Blanc Bleu Belge	(Fasquelle <i>et al.</i> , 2009) (Sartelet <i>et al.</i> , 2012)

<i>PITX3</i>	Microphthalmie	OAR22 : c.338G>C	Texel	(Becker <i>et al.</i> , 2010)
<i>RNASEH2B</i>	Mortalité embryonnaire	BTA12 : g.20100649del662463bp	Nordic Red	(Kadri <i>et al.</i> , 2014)
<i>SDE2</i>		BTA16 : g.29773628A>G	Holstein	(Fritz <i>et al.</i> , 2018)
<i>SHBG</i>		BTA19 : p.Q52X	Montbéliarde	(Fritz <i>et al.</i> , 2013)
<i>SLC35A3</i>	CVM	BTA3 : p.V180F	Holstein	(Thomsen <i>et al.</i> , 2006)
<i>SLC37A2</i>	Mortalité embryonnaire	BTA7 : p.R12X	Normande	(Fritz <i>et al.</i> , 2013)
<i>SLC6A5</i>	Dystonie	BTA29 : p.L270P	Blanc Bleu Belge	(Charlier <i>et al.</i> , 2008)

Abréviations : ABCA12 : Atp Binding Cassette subfamily A member 12 ; ABCG2 : Atp Binding Cassette subfamily G member 2 ; APAF1 : Apoptotic Peptidase Activating Factor 1 ; APOB : ApoliPOprotein B ; B4GALNT ; beta-1,4-N-acetyl-galactosaminyl transferase ; BCO2 : beta-carotene oxygenase 2 ; BMP : Bone Morphogenetic Protein (protéine morphogénétique osseuse) ; BMPR : Bone Morphogenetic Protein Receptor (récepteur de protéines morphogénétiques osseuses) ; BTA : Bos TAurus ; CAD : Carbamoyl-phosphate synthetase 2, Aspartate transcarbamylase, and Dihydroorotase ; CHI : Capra HIrcus ; CSN : Casein ; CVM : Complex Vertebral Malformation (malformation de la colonne vertébrale) ; DAL : Défaut d'Adhérence des Leucocytes ; DGAT : DiacyGlycerol o-AcylTransferase ; FANCI : Fanconi ANemia Complementation group 1 ; FEZL : Forebrain Embryonic Zinc finger-Like ; GDF : Growth Differentiation Factor (facteur de différenciation de la croissance) ; ITGB : InTeGrin subunit Beta ; MITF : Melanocyte Inducing Transcription Factor ; MRC2 : Mannose Receptor C type 2 ; MSTN : Myostatin ; OAR : Ovis ARies ; PITX3 : Paired Like Homeodomain 3 ; PRNP : PRioN Protein (protéine prion) ; RNASEH2B : RiboNucleASE H2 subunit B ; SDE2 : SDE2 telomere maintenance homolog ; SLC35A3 : SoLute Carrier family 35 member A3 ; SLC37A2 : SoLute Carrier family 37 member A2 ; SOCS2 : Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines) ; TMEM : Transmembrane Protein (protéine transmembranaire) ; TYRP1 : TYrosinase-Related protein 1.

Les caractères liés à la reproduction sont également peu déterminés par des gènes majeurs (Tableau 1). Les principaux décrits dans la littérature sont dans l'espèce ovine où quatre gènes sont identifiés comme déterminant le taux d'ovulation chez la brebis (Abdoli *et al.*, 2016). Ces gènes sont aussi de grand intérêt pour les filières allaitantes car ils déterminent la prolificité, un caractère également lié à la production. En bovin, un seul gène majeur sur la fertilité est à notre connaissance mis en évidence, sans mutation causale pour le moment découverte (Kamalludin *et al.*, 2018), et un QTL est trouvé par plusieurs équipes indépendantes de chercheurs sur le chromosome 18 (Ma *et al.*, 2019).

Pour finir, de nombreux variants/gènes (Tableau 1) responsables d'anomalies génétiques ont été identifiés ces dernières années. Ces variants sont non souhaitables dans les populations car causant des malformations ou des mortalités embryonnaires et/ou précoces des individus, et représentant une perte économique. Ces anomalies sont principalement trouvées grâce à des analyses visant à détecter les *loci* où des déficits en homozygotes sont présents dans les populations, c'est-à-dire qu'il n'existe pas d'individu homozygote mutant pour un *locus* polymorphe donné. De par la précision de ce phénotype de déficit en homozygote associé à une malformation, la possibilité de trouver un gène majeur le déterminant est grande. Ces recherches sont très avancées en bovin. En outre, des anomalies génétiques supplémentaires chez le bovin allaitant, non présentes dans le Tableau 1, ont été rapportées récemment dans un article de synthèse bibliographique (Cieplóch *et al.*, 2017).

I.3.b. Quelques exemples de prise en compte de mutations dans les schémas de sélection

Dans les schémas de sélection, les gènes majeurs sont aujourd'hui pris en compte indépendamment des index synthétiques proposés aux éleveurs. C'est le cas par exemple des caséines en caprin où les boucs sont présélectionnés au sein des centres de testage sur leur allèle de la caséine alpha-s1 (Piacère *et al.*, 1997), les allèles forts, associés à un plus fort taux protéique (meilleure qualité technologique du lait), étant préférentiellement disséminés au sein de la population commerciale. Des études récentes ont cherché à inclure l'information de ce gène majeur dans les modèles d'évaluation (Carillier-Jacquin *et al.*, 2016 ; Teissier *et al.*, 2018), mais les solutions prometteuses proposées ne sont pas encore mises en place en routine dans l'évaluation génomique caprine actuelle.

Les gènes majeurs associés à des maladies spécifiques sont également sélectionnés indépendamment des index. Par exemple, la tremblante (maladie létale à prion de la famille des Encéphalopathies Spongiformes Subaiguës Transmissibles touchant le système nerveux central

et sans traitement connu à ce jour) a été associée à un gène dont trois mutations causales provoquent l'apparition d'haplotypes (combinaison d'acides aminés modifiés au niveau des trois *loci* où les mutations causales sont présentes) plus ou moins sensibles : VRQ, ARQ, AHQ, et ARR. C'est sur la base de ces quatre haplotypes qu'est réalisée la sélection en France depuis 2002 au sein du Programme National d'Amélioration Génétique pour la Résistance à la Tremblante (Palhière *et al.*, 2008 ; Tortereau, 2016) dans les deux filières ovines laitière et allaitante. L'objectif de cette sélection était d'éliminer l'allèle « sensible » VRQ dans la population commerciale en disséminant des béliers porteurs de l'allèle « résistant » ARR. Le résultat de cette sélection est aujourd'hui une quasi fixation de l'allèle ARR dans toutes les populations sélectionnées, avec une faible augmentation de la consanguinité grâce à la bonne gestion de la variabilité génétique dans le cadre du programme national.

Les anomalies génétiques récessives sont encore aujourd'hui peu sélectionnées dans les schémas, les cas d'homozygotes vivants étant assez rares et les phénotypes non définis comme des caractères d'intérêt majeur par les filières. Ces anomalies sont aujourd'hui assez répandues dans les populations commerciales en raison bien souvent d'une dissémination à grande échelle d'un individu améliorateur porteur de l'anomalie à l'état hétérozygote, comme c'est le cas par exemple pour le syndrome Brachyspina pour lequel tous les cas rapportés remontent à un taureau autrefois très populaire *Sweet Haven Tradition* (Charlier *et al.*, 2012). Les mâles reproducteurs porteurs de ces anomalies étaient jusqu'à présent simplement retirés des schémas lorsqu'ils étaient détectés. Avec le développement des techniques de génotypage et la découverte de nombreuses anomalies ces vingt dernières années chez les bovins, il serait possible d'ajouter les génotypes de ces *loci* identifiés au sein des index, avec une pondération économique liée à l'anomalie préalablement estimée, comme le suggère Hoze et ses co-auteurs dans le schéma des bovins laitiers en France (Hoze *et al.*, 2018). Cette dernière approche considère toutefois que les mutations sont indépendantes les unes des autres et qu'elles n'influencent pas les caractères de l'objectif de sélection (corrélation génétique nulle).

A notre connaissance, bien que rares à ce jour, il existe des exemples de gènes majeurs à effets pléiotropes chez les ruminants. C'est le cas par exemple du syndrome d'intersexualité sans-corne PIS (*Polled Intersex Syndrome*), présent chez les caprins, et dû à une délétion de 11,7 kilobases sur le chromosome 1 (Pailhoux *et al.*, 2001 ; Pailhoux *et al.*, 2002 ; Boulanger *et al.*, 2014), impactant notamment le gène *FOXL2* (*Forkhead Box L2*), essentiel pour la différenciation ovarienne. Les individus de type sauvage sont cornus et normalement sexués. Lorsque la mutation est présente à l'état hétérozygote, les individus sont sans cornes et normalement sexués. En revanche, les individus homozygotes pour la mutation sont sans cornes

et sexués pour les mâles (avec des cas de stérilité selon l'état d'obstruction de l'épididyme), et intersexuées pour les femelles (toutes stériles, dû à une modification du développement précoce des gonades). Ce cas illustre bien la pléiotropie d'un gène impacté par une mutation sur la détermination d'un phénotype de morphologie fonctionnelle (corne) et d'un phénotype de reproduction.

Parmi les tares génétiques décrites dans le Tableau 1, le gène *MRC2* (*Mannose Receptor C type 2*) (Fasquelle *et al.*, 2009) présente un effet pléiotrope sur la morphologie et la muscularité en race bovine Blanc Bleu Belge. Ce gène est en effet impliqué dans la régulation de la dégradation et du remodelage de la matrice extracellulaire et est fortement exprimé dans les os en développement. En 2008, une première mutation, associée au syndrome de la queue tordue, et expliquant 3,6% de la variance génétique de la muscularité, a été identifiée dans ce gène (délétion de deux nucléotides dans l'exon 20). Cette mutation est associée à des malformations du squelette (tête raccourcie, développement retardé) et des muscles (hypertrophie très marquée principalement au niveau du muscle *gluteus medius* de l'arrière-train). La fréquence de cette mutation dans la population commerciale était de 25% en 2008 en raison d'une forte dissémination du fondateur porteur, le taureau Précieux, né en 1980. Pour la gestion de cette mutation, en 2008, un test de diagnostic a été préparé et appliqué chez les taureaux d'IA pour réduire sa fréquence et limiter les pertes économiques au sein des élevages (25% des animaux homozygotes porteurs de la mutation doivent être euthanasiés et les 75% restant ont un retard de croissance prononcé et la qualité de la carcasse est diminuée). Ainsi, l'incidence de ce syndrome a disparu en 2009 mais de nouveaux cas ont été découverts fin 2009 et début 2010 et ont mis en lumière une seconde mutation dans ce gène provoquant les mêmes symptômes dus à la protéine correspondante non fonctionnelle (Sartelet *et al.*, 2012). Cet exemple montre la problématique des gènes pléiotropes dans le cadre de la sélection. En effet, favorable pour les caractères musculaires à l'état hétérozygote, les deux mutations dans le gène *MRC2* sont fortement défavorables à l'état homozygote pour d'autres caractères morphologiques, pouvant aller même jusqu'à l'obligation pour l'éleveur d'euthanasier l'animal pour des raisons de bien-être. Les mutations de ce gène pléiotrope sont non désirables au vu de leurs conséquences dramatiques si bien que les acteurs de la filière souhaitent les éradiquer. Dans le cas que nous allons décrire dans la section suivante, les effets pléiotropes sont moins tranchés et la question de la gestion de la mutation semble être plus difficile à traiter.

1.3.c. Cas particulier de la mutation R96C du gène *SOCS2*

La mutation à l'étude dans cette thèse a été découverte associée aux mammites dans une population de brebis laitières (Rupp *et al.*, 2015). En effet, à l'aide de GWAS sur une puce à

SNP de moyenne densité (50 000 marqueurs) pour le caractère associé aux mammites CCS, un QTL sur le chromosome 3 ovin a été mis en évidence aux alentours de 130 Mb chez la brebis de race Lacaune.

Cette région a ensuite été cartographiée finement par les auteurs grâce au séquençage complet de trois animaux : un père possédant l'haplotype supposé hétérozygote dans la région détectée et deux de ses fils supposés homozygotes (type sauvage et mutant). Cette cartographie fine a fourni un SNP candidat qui correspondait à une base nucléotidique C modifiée en T en position 129 722 200 paires de bases sur le chromosome 3 de l'assemblage v3.1 ovin (ftp://ftp.ensembl.org/pub/release-97/fasta/ovis_aries/dna/). Cette substitution se situe dans la séquence codante du gène hautement conservé *SOCS2* (Figure 4), et plus particulièrement dans son domaine fonctionnel SH2 (*Src Homology 2* - domaine 2 homologue à la protéine Src), site de liaison de la protéine à divers ligands, et provoque une modification d'un acide aminé dans la protéine correspondante (Arginine - R - modifiée en Cystéine - C -, en position 96 de la protéine).

A l'aide d'un test fonctionnel, les auteurs (Rupp *et al.*, 2015) ont montré que la mutation ponctuelle p.R96C annule complètement l'affinité de liaison de la protéine SOCS-2 pour son ligand principal connu dans la littérature, la phosphotyrosine pY595 du récepteur à l'hormone de croissance, suggérant ainsi la perte de l'activité fonctionnelle de la protéine.

Le génotype de cette mutation *SOCS2* explique 12% de la variance du caractère dans une population de 468 mâles améliorateurs et la fréquence de l'allèle mutant dans cette même population était de 21,7%, dont seulement 1,5% d'homozygotes mutés. Afin d'expliquer la haute fréquence de cette mutation désavantageuse pour la santé, les auteurs ont cherché des associations entre cette mutation et des caractères de production. Ils ont ainsi pu mettre en évidence une augmentation de 24%, 18% et 4,4% respectivement de la taille, du poids et de la production laitière entre des ovins homozygotes p.R96C par rapport aux ovins sauvages. Cette mutation ayant un impact positif sur la production laitière, il est possible qu'elle ait pu être sélectionnée intentionnellement ou non au cours des générations, malgré son impact négatif sur la santé mammaire.

Depuis cette découverte, il est recommandé aux OS de ne pas conserver parmi leurs mâles améliorateurs les individus homozygotes pour cette mutation en raison de son très fort impact négatif sur le Score de Cellules Somatiques (SCS) dans le lait des brebis. En effet, en plus du bien-être animal qui peut être mis à mal en raison de mammites chroniques associées à cette

mutation, un impact économique négatif pour les éleveurs est réel, dû au paiement du lait réduit pour cause d'un très grand nombre de cellules dans le lait commercialisé. Le devenir de cette mutation n'est pour le moment non encore statué entre les différents acteurs de la filière en France.

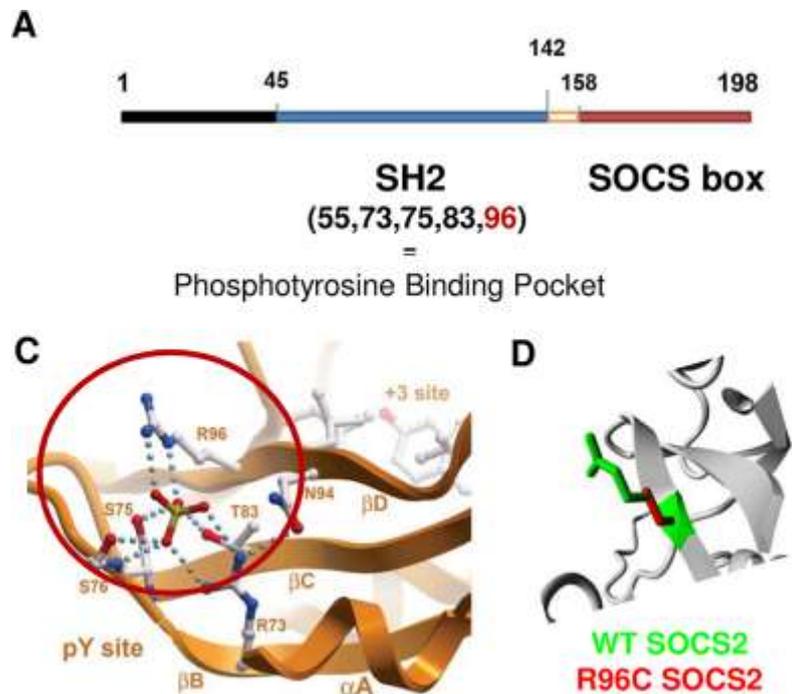


Figure 4 Caractérisation bioinformatique de la mutation R96C dans le gène ovin SOCS2 (Rupp et al., 2015)

SH2 : Src Homology 2 (domaine 2 homologue à la protéine Src) ; SOCS : Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines) ; WT : Wild Type (type sauvage) ; (A) Structure du gène SOCS2 (C) Site de la mutation dans la structure de la protéine (D) Zoom sur la mutation. La protéine est de couleur grise, les chaînes latérales du type sauvage et du résidu mutant sont représentées respectivement en vert et en rouge.

II. SOCS-2, une protéine impliquée dans de nombreuses fonctions biologiques

Le gène *SOCS2* code pour une protéine (SOCS-2) qui fait partie d'une famille de huit protéines régulatrices des signaux de transduction du signal, les protéines SOCS. Le principal rôle connu des protéines SOCS est de réguler une voie majeure de transmission du signal, la voie de signalisation JAK (*Janus Kinase*) / STAT (*Signal Transducer and Activator of Transcription* - transducteur de signal et activateur de transcription). Cette voie permet la transformation du signal extracellulaire en un signal intracellulaire. La transmission du signal *via* la voie JAK/STAT s'effectue grâce à la fixation de molécules de signalisation (hormones) sur leurs récepteurs membranaires qui déclenche une cascade de phosphorylations. Cette cascade est régulée de manière complexe, et aboutit à la transcription de gènes spécifiques en réponse à la fixation du ligand au récepteur. Ces protéines SOCS ont donc souvent un effet pléiotrope, étant donné qu'elles régulent de nombreuses fonctions biologiques selon la voie de signalisation ciblée.

Dans cette partie, nous décrirons brièvement les mécanismes généraux de cette voie de signalisation JAK/STAT, ainsi que les principales molécules activatrices et régulatrices de cette voie. Puis, nous décrirons les protéines SOCS et les fonctions de régulation de cette voie. Enfin, un inventaire des différentes fonctions régulées par la protéine SOCS-2 en particulier sera effectué.

II.1. La voie de transmission du signal cellulaire JAK/STAT

La voie JAK/STAT a été initialement identifiée comme étant la voie de transmission du signal à partir des récepteurs à de nombreuses cytokines et hormones polypeptidiques (Kishimoto *et al.*, 1994) produites en réponse à différents *stimuli*. Elles interviennent dans la régulation de plusieurs fonctions biologiques comme la prolifération et la différenciation cellulaires ou encore la réponse immunitaire. Cette voie de transduction du signal depuis les récepteurs de cytokines a ensuite été identifiée comme partagée avec plusieurs facteurs de croissance tels que l'hormone de croissance ou encore la prolactine. Il existe en effet une grande similarité entre les récepteurs membranaires de l'ensemble de ces molécules de signalisation (Cosman *et al.*, 1990). La voie JAK/STAT est donc impliquée dans la transmission de signaux régulant de nombreuses fonctions biologiques.

II.1.a. Principe général de cette voie de transmission

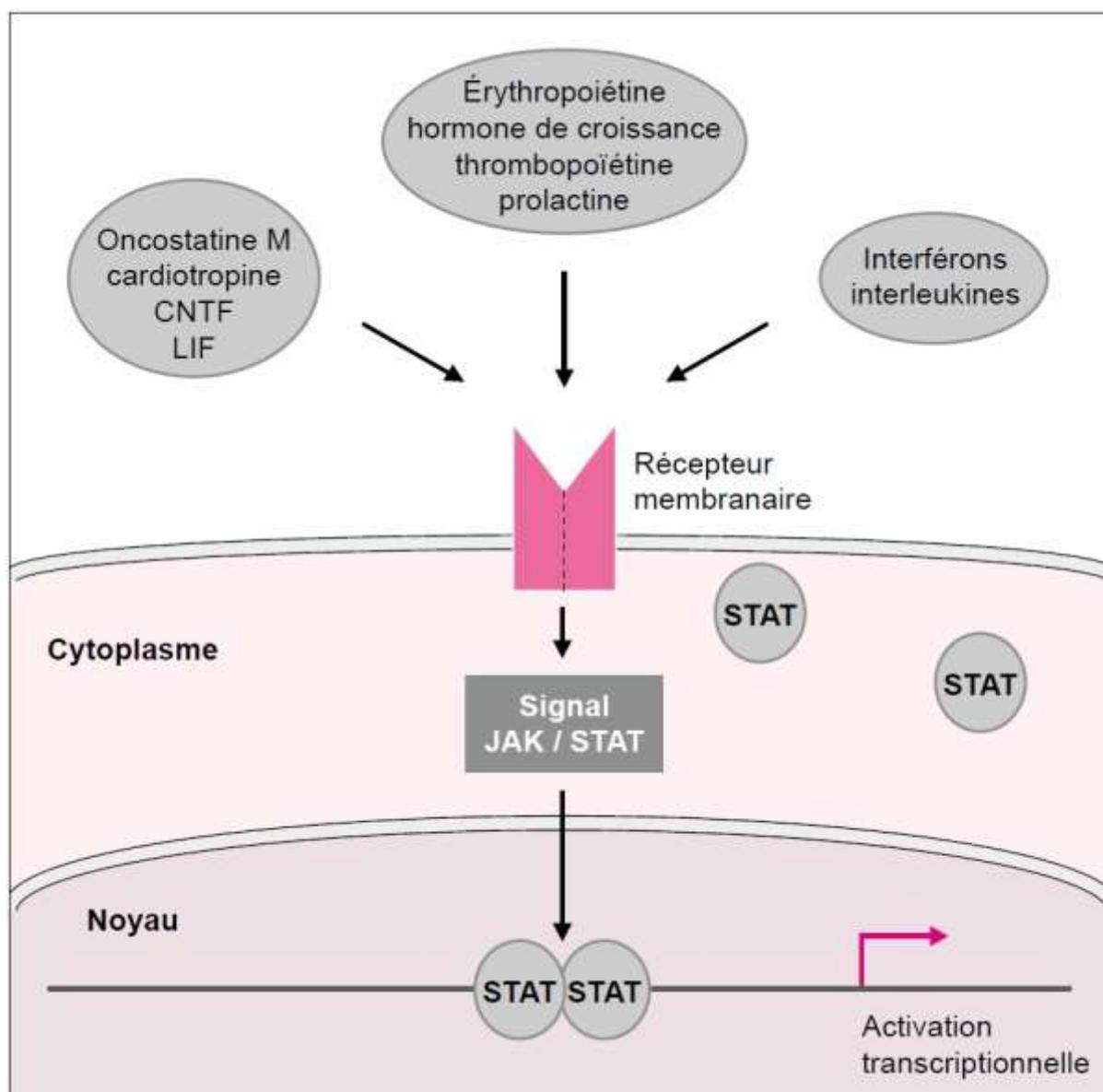


Figure 5 Principe général de la voie de transmission du signal JAK/STAT (Vignais, 1997)

CNTF : Ciliary Neurotrophic Factor (facteur neurotrophique ciliaire) ; JAK (Janus Kinase) ; LIF : Leukemia Inhibitory Factor (facteur inhibiteur de la leucémie) ; STAT : Signal Transducer and Activator of Transcription (transducteur de signal et activateur de transcription)

La transmission du signal *via* la voie JAK/STAT (Igaz *et al.*, 2001 ; Ivashkiv & Hu, 2004 ; Rawlings *et al.*, 2004 ; Schindler & Plumlee, 2008 ; Mohr *et al.*, 2012) débute avec l'activation de récepteurs membranaires après liaison de différentes cytokines et hormones évoquées précédemment (Figure 5). Les protéines JAK, qui se lient aux chaînes intracellulaires du récepteur de manière constitutive, vont se phosphoryler mutuellement et induire le recrutement de deux protéines STAT qui sont, à ce stade, libres dans le cytoplasme. Les protéines STAT se fixent au récepteur, pour être à leur tour phosphorylées par les protéines JAK et vont se dimériser. Ces dimères STAT-STAT activés vont alors pouvoir migrer dans le noyau et se fixer

à une séquence ADN afin d'activer la transcription d'un ensemble de gènes spécifiques de la molécule initialement fixée au récepteur membranaire.

II.1.b. Les protéines JAK

Découvertes dans la même période que la voie de signalisation des STAT (Firmbach-Kraft *et al.*, 1990 ; Wilks *et al.*, 1991), quatre protéines de la famille JAK sont à ce jour connues et décrites chez les Mammifères (Murray, 2007 ; Schindler & Plumlee, 2008 ; Ghoreschi *et al.*, 2009 ; Ferrao & Lupardus, 2017 ; Villarino *et al.*, 2017) : JAK-1, -2, -3, et Tyk2 (*Tyrosine kinase 2*).

Ces quatre protéines possèdent une structure commune composée de sept domaines homologues, JH (*Jak Homology*) 1 à 7 (Figure 6). Parmi ces domaines, seul JH1 contient les résidus tyrosine nécessaires à l'activité de kinase (domaine Ki). Le domaine pseudo-kinase (Ψ Ki), qui est une portion de JH2, sert à limiter les activités de kinase non justifiées grâce à sa capacité de réguler négativement l'activité de la protéine JAK en phosphorylant ses sites catalytiques. A cheval entre JH3 et JH4 se trouve le domaine de type SH2 qui intervient dans l'interaction avec les récepteurs en amont de la fixation du ligand. Enfin, les domaines JH4 à 7 constituent le domaine appelé FERM (*Four-point-one, Ezrin, Radixin, Moesin* - domaine homologue à la protéine quatre-point-un) qui intervient dans l'interaction avec les récepteurs et favorise la fonction kinase.

Lors de la fixation du ligand à son récepteur, les protéines JAK fixées au récepteur se rapprochent et s'activent par phosphorylation croisée. Bien que le nombre de ligands en capacité de se fixer à un récepteur soit élevé, il existe un nombre restreint de combinaisons possibles pour la formation de dimères de protéines JAK (homodimères ou hétérodimères) (Figure 7). Les combinaisons bien caractérisées dans la littérature sont les homodimères JAK-1/JAK-1 et JAK-2/JAK-2, et l'hétéro-dimère JAK-1/JAK-3. Par exemple, le récepteur à l'hormone de croissance (GH : *Growth Hormone*) est complexé avec deux protéines JAK-2 qui s'y fixent grâce à l'extrémité N-terminale de leur domaine FERM (Argetsinger *et al.*, 1993).

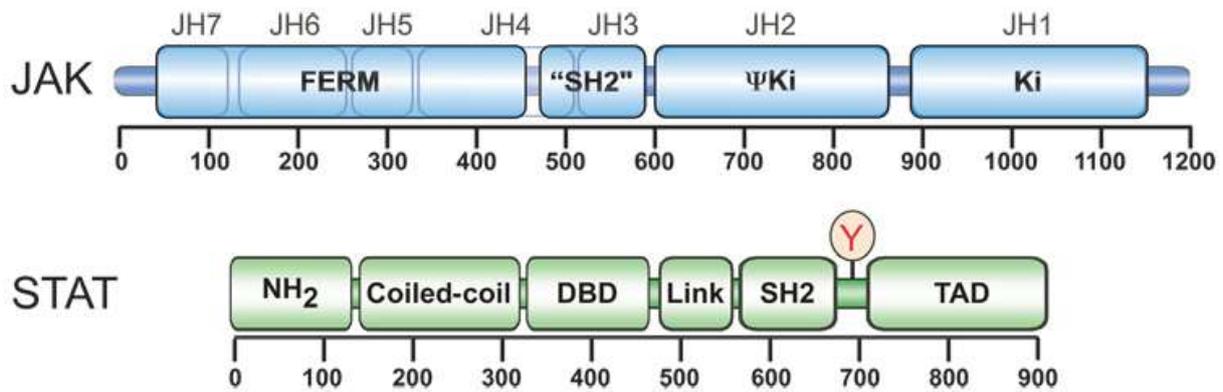


Figure 6 Organisation structurale commune aux différentes protéines de chacune des familles JAK et STAT (Schindler & Plumlee, 2008)

DBD : Dna Binding Domain (domaine de liaison à l'ADN) ; FERM : Four-point-one, Ezrin, Radixin, Moesin (domaine homologue à la protéine 4.1) ; JAK (Janus Kinase) ; JH : Jak Homology domain (domaine d'homologie JAK) ; Ki : Kinase ; NH2 : amino terminal domain (domaine N-terminus) ; ΨKi ; Pseudo Kinase ; STAT : Signal Transducer and Activator of Transcription (transducteur de signal et activateur de transcription) ; SH2 : Src Homology 2 (domaine 2 homologue à la protéine Src) ; TAD : Transcriptional Activation Domain (domaine à activité transcriptionnelle)

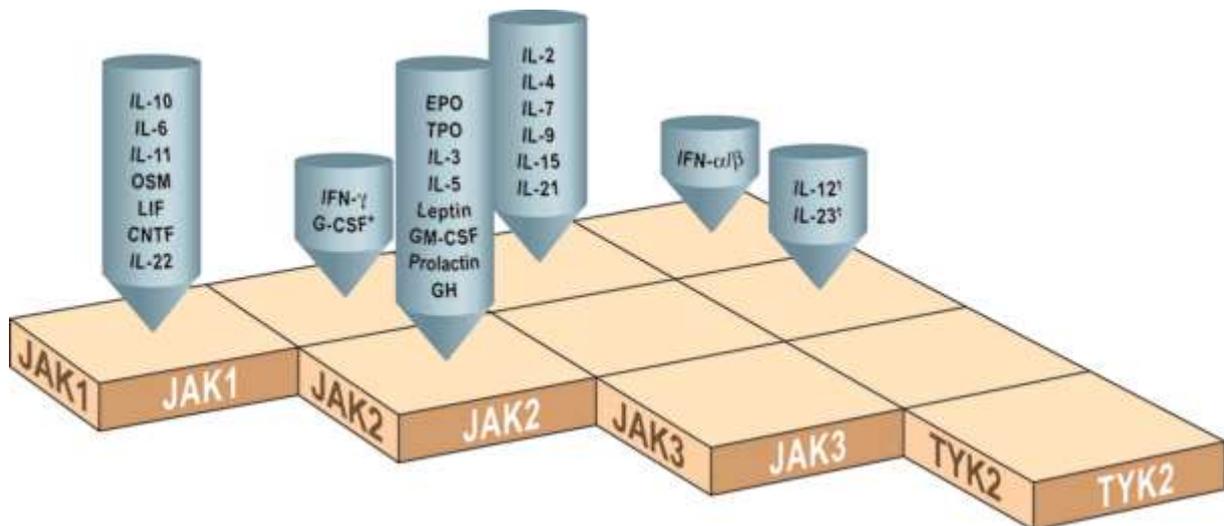


Figure 7 Combinaisons connues des protéines JAK intervenant dans la transmission du signal des principales molécules utilisant la voie JAK/STAT (Murray, 2007)

* † : Combinaisons non confirmées ; CNTF : Ciliary Neurotrophic Factor (facteur neurotrophique ciliaire) ; G-CSF : Granulocyte-Colony Stimulating Factor (facteur de stimulation des colonies de granulocytes) ; GM-CSF : Granulocyte Macrophage-Colony-Stimulating Factor (facteur de stimulation des colonies de granulocytes et de macrophages) ; IL : InterLeukin (interleukine) ; IFN : InterFeroN (interféron) ; JAK (Janus Kinase) ; LIF : Leukemia Inhibitory Factor (facteur inhibiteur de la leucémie) ; OSM : OncoStatin-M (oncostatine M)

II.1.c. Les protéines STAT

Cette famille de protéines, découverte également à la fin des années 80 (Kessler *et al.*, 1988a ; Kessler *et al.*, 1988b ; Levy & Darnell, 2002), est composée de sept facteurs de transcription partageant une structure commune chez les mammifères : STAT-1, -2, -3, -4, -5a, -5b et -6 (Rawlings *et al.*, 2004 ; Schindler & Plumlee, 2008 ; Wieczorek *et al.*, 2012).

Les protéines STAT contiennent sept domaines fonctionnels conservés (Figure 6) :

- 1) Le domaine NH2 (extrémité N-terminale) a pour fonction de favoriser la formation de dimères parmi les STAT non phosphorylées, ce qui permettrait non seulement de s'assurer que les STAT restent dans une conformation non active à l'état de repos dans le cytoplasme, mais faciliterait également la fixation et l'activation ultérieure des dimères au niveau des récepteurs.
- 2) Le domaine « *Coiled-coil* » est un domaine en super-hélice permettant le recrutement de protéines dans des complexes de régulation transcriptionnelle ; il est également impliqué dans le transport nucléaire de ces complexes
- 3) Le domaine DBD (*Dna Binding Domain*) est le domaine de liaison à l'ADN. Ce domaine possède également un rôle de localisation des homodimères.
- 4) Le domaine Link est un domaine d'interaction protéique permettant la liaison entre protéines STAT activées.
- 5) Le domaine SH2 (domaine le plus conservé) permet la liaison de la protéine STAT au complexe récepteur/JAK.
- 6) Le domaine suivant (Y) est la localisation du résidu tyrosine qui doit être phosphorylé par les protéines JAK.
- 7) Enfin, le domaine TAD (*Transcriptional Activation Domain* - domaine à activité transcriptionnelle) est très variable selon la protéine STAT et peut contenir également un site de phosphorylation (sérine) qui facilite le recrutement de cofacteurs. Il serait également impliqué dans la stabilité de la protéine (dégradation protéasomique).

Les différentes protéines STAT non activées sont libres dans le cytoplasme sous diverses formes (monomères, homodimères et hétérodimères). Lors de la fixation du ligand sur le récepteur, une phosphorylation des JAK est réalisée, permettant la libération de sites d'ancrage pour les STAT qui viennent alors se fixer au complexe ligand/récepteur/JAK et ainsi entraîner la cascade de signalisation cellulaire. Pour illustrer cela, la protéine STAT-5b joue un rôle majeur dans la signalisation de l'hormone de croissance (Udy *et al.*, 1997). Elle vient se fixer au récepteur de la GH *via* son domaine SH2 et sa phosphorylation par les protéines JAK-2, à certaines localisations (Ser193, Tyr699, et PSP731), est nécessaire pour son activité transcriptionnelle suite à sa migration dans le noyau (Wójcik *et al.*, 2018).

II.1.d. Autres molécules intervenant dans la voie de transmission JAK/STAT

Cette voie de transmission est complexe et finement régulée par un grand nombre de molécules différentes. Des protéines adaptatrices sont identifiées dans la littérature comme pouvant intervenir dans la régulation de la voie JAK/STAT (Igaz *et al.*, 2001 ; Shuai & Liu, 2003 ; Rawlings *et al.*, 2004 ; Schindler & Plumlee, 2008). En voici une liste non exhaustive :

- les protéines de la famille des STAM (*Signal-Transducing Adapter Molecule* - molécule adaptatrice à transduction de signal) sont des molécules qui peuvent faciliter l'activation de la transcription de certains gènes par un mécanisme mal connu faisant intervenir leur phosphorylation par les protéines JAK (Lohi & Lehto, 2001).
- les protéines StIP (*Stat-Interacting Protein* - protéine interagissant avec les STAT) facilitent la phosphorylation des STAT par les JAK grâce à leur capacité de se fixer aux deux molécules (Collum *et al.*, 2000).
- les molécules des familles SH2B/Lnk/APS sont des substrats des protéines JAK grâce à leurs domaines SH2 et homologues de la pleckstrine (O'Brien *et al.*, 2002). Ces molécules peuvent soit inhiber (Wakioka *et al.*, 1999), soit faciliter (Maures *et al.*, 2007) la voie de transmission JAK/STAT.

Trois grandes classes d'inhibiteurs de la voie JAK/STAT ont été décrites (Greenhalgh & Hilton, 2001 ; Wormald & Hilton, 2004 ; Villarino *et al.*, 2017) (Figure 8) : les phosphatases de la famille PTP (*Protein Tyrosine Phosphatases* - protéines tyrosine phosphatases), et les familles de protéines PIAS (*Protein Inhibitors of Activated Stat* - protéines inhibitrices des STAT actives) et SOCS.

- Les SOCS (1) sont des protéines dont la transcription est directement induite par les STAT et qui peuvent se fixer *via* leur domaine SH2 soit aux récepteurs, soit aux JAK pour entrer en compétition avec les protéines STAT.
- Les PTP (2) interviennent dans la voie JAK/STAT grâce à leur capacité à déphosphoryler les phosphorylations effectuées par les protéines JAK sur les protéines STAT (Böhmer & Friedrich, 2014).
- Les PIAS (3) quant à elles ont la capacité de se lier aux STAT dimérisées et d'empêcher leur interaction avec l'ADN (Chung *et al.*, 1997).

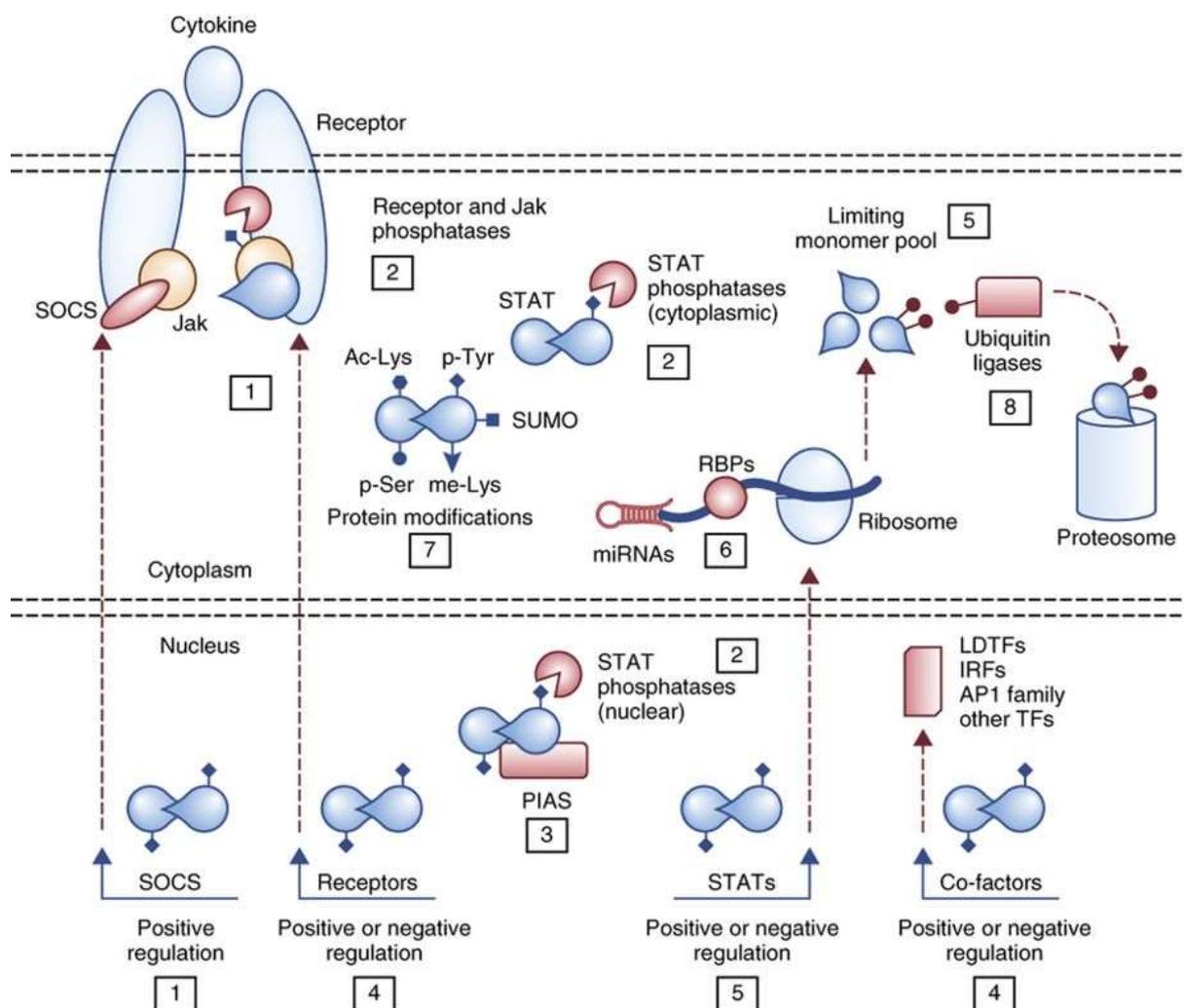


Figure 8 Les nombreux niveaux de régulation de la voie JAK/STAT (Villarino *et al.*, 2017)
 JAK : Janus Kinase ; LDTFs, IRFs, AP1 et TF : facteurs de transcription ; PIAS : Protein Inhibitors of Activated Stat (protéines inhibitrices des STAT actives) ; RBP : RNA-binding protein (protéine de liaison à l'ARN) ; SOCS : Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines) ; STAT : Signal Transducer and Activator of Transcription (transducteur de signal et activateur de transcription) ; SUMO : Small Ubiquitin-Related Modifier (petit modificateur lié à l'ubiquitine)

Enfin, des niveaux plus fins de régulation des STAT existent (Figure 8), tels que :

- La régulation transcriptionnelle qui est directement effectuée par les STAT grâce à leur capacité de liaison à l'ADN, activant la transcription de gènes codant pour les récepteurs et des facteurs de transcription (cofacteurs) qui influencent l'activité des STAT (4), ainsi que de gènes codant les STAT eux-mêmes (5) (Cheon *et al.*, 2011).
- Les microARN (ou miARN) et les protéines de liaison à l'ARN, qui influencent la demi-vie et la traduction de l'ARN messager des STAT (6) (Kohanbash & Okada, 2012).
- Les modifications post-traductionnelles : phosphorylation de la sérine, acétylation, méthylation et sumoylation de la lysine qui peuvent favoriser ou supprimer l'activité des STAT (7) (Wieczorek *et al.*, 2012).

- Les ubiquitines ligases qui sont responsables de la dégradation protéasomique des STAT (8) (Kim & Maniatis, 1996).

Parmi tous ces mécanismes de régulation, nous allons nous intéresser à une famille d'inhibiteurs dont fait partie la protéine au cœur de l'étude de cette thèse (SOCS-2), la famille de protéines SOCS.

II.2. Les protéines SOCS

II.2.a. Structure caractéristique et fonctions des protéines SOCS

Cette famille de protéines inhibitrices de la voie de signalisation JAK/STAT a été découverte par plusieurs équipes de chercheurs à la fin des années 90 (Yoshimura *et al.*, 1995 ; Endo *et al.*, 1997 ; Matsumoto *et al.*, 1997 ; Naka *et al.*, 1997 ; Starr *et al.*, 1997).

On compte aujourd'hui chez les mammifères huit protéines SOCS possédant une structure commune conservée : CIS, SOCS-1 à -7 (Figure 9) (Alexander, 2002 ; Fujimoto & Naka, 2003 ; Croker *et al.*, 2008 ; Villarino *et al.*, 2017 ; Yoshimura *et al.*, 2018).

Ces huit protéines possèdent une structure commune composée d'une région centrale SH2 qui permet la liaison à un substrat (récepteur membranaire ou protéine JAK selon la protéine SOCS) et d'une boîte SOCS (*SOCS-box*) qui a la capacité de se lier à un complexe d'ubiquitines (polyubiquitination du complexe protéique en vue de sa dégradation par le protéasome) (Coux & Piechaczyk, 2000). Une région KIR (*Kinase Inhibitory Region* - région inhibitrice de kinase, en rouge sur la Figure 9), qui peut directement interagir avec les protéines JAK, est présente dans SOCS-1 et SOCS-3.

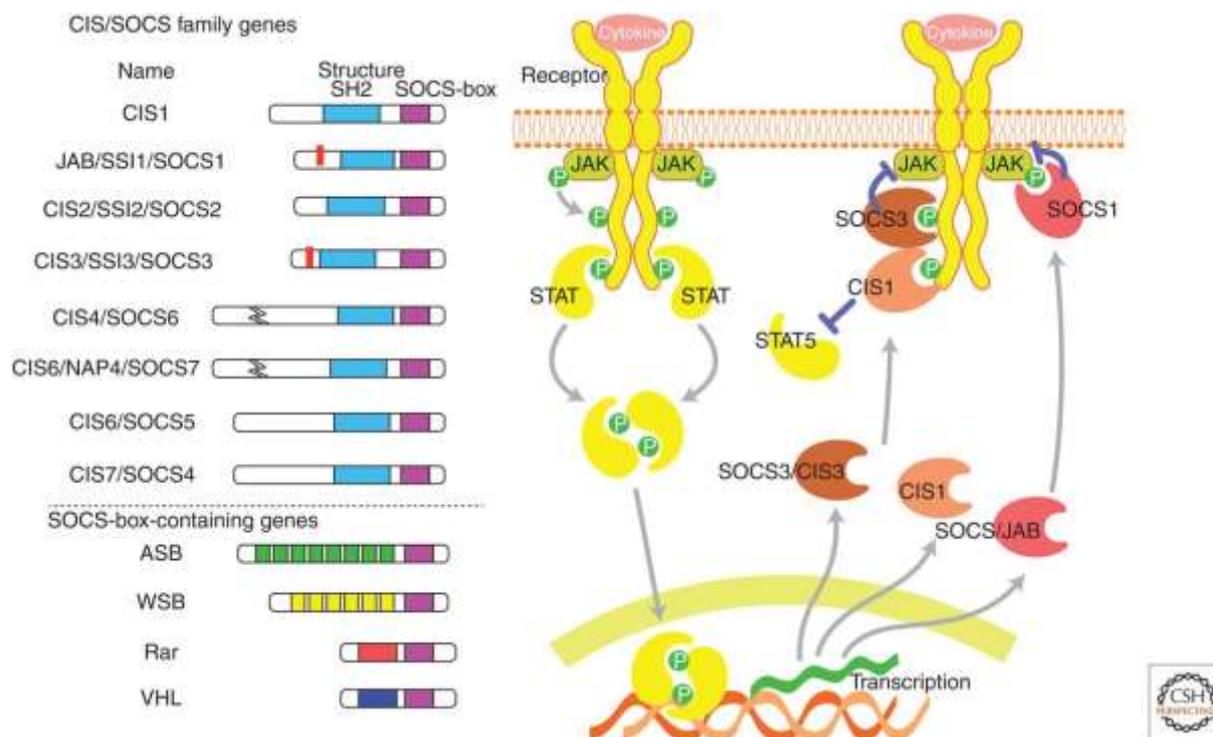


Figure 9 Structure et fonction de la famille des protéines SOCS dans la voie de transmission JAK/STAT (Yoshimura *et al.*, 2018)

ASB : Ankyrin repeat and SOCS Box ; CIS : Cytokine-Inducible SH2 domain-containing protein (protéine contenant un domaine SH2 inducible aux cytokines) ; JAB : Jak Binding protein (protéine se liant à JAK) ; JAK : Janus Kinase ; NAP4 : NCK(Non-Catalytic region of tyrosine Kinase)-Associated Protein 4 (protéine associée à la région non catalytique de la tyrosine kinase 4) ; Rar : RAs Related ; SH2 : Src Homology 2 (domaine 2 homologue à la protéine Src) ; SOCS : Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines) ; SSI : Stat-induced Stat Inhibitor (inhibiteur de STAT induit par STAT) ; STAT : Signal Transducer and Activator of Transcription (transducteur de signal et activateur de transcription) ; VHL : Von Hippel–Lindau ; WSB : WD40 repeat and SOCS box

Comme l'indique la Figure 9, les protéines SOCS sont produites en réponse, entre autres, au signal des STAT dimérisées qui ont migré dans le noyau. Ces protéines participent donc à la régulation négative de la voie de transmission JAK/STAT. Leur régulation peut s'effectuer selon trois mécanismes hypothétiques :

- 1) La protéine SOCS peut se fixer *via* son domaine SH2 soit au récepteur membranaire (Qing *et al.*, 2005), soit aux JAK (Yasukawa *et al.*, 1999), et ainsi empêcher, par compétition, l'activation des facteurs STAT, possédant le même domaine de fixation, qui resteront alors inactifs dans le cytoplasme.
- 2) La boîte SOCS peut recruter les élongines B et C qui vont permettre l'ubiquitination des complexes protéiques liés à SOCS et ainsi conduire à leur dégradation par le protéasome (Hilton *et al.*, 1998 ; Zhang *et al.*, 1999). Cette dégradation provoque l'arrêt de transmission du signal. La boîte SOCS permet également l'interaction avec le récepteur, la stabilité de la protéine SOCS elle-même, ainsi que la régulation des autres protéines SOCS dans certains cas (Piessevaux *et al.*, 2008).

- 3) Les protéines SOCS-1 et -3 peuvent également inhiber l'activité kinase des JAK grâce à leur région KIR (Yasukawa *et al.*, 1999), empêchant ainsi l'activation des STAT, leur dimérisation et la transmission du signal.

De nombreuses études se sont penchées sur les multiples molécules de signalisation régulées par les protéines SOCS au sein de la voie JAK/STAT. Le tableau suivant (Tableau 2), inspiré de trois articles de synthèse bibliographiques (Alexander, 2002 ; Fujimoto & Naka, 2003 ; Croker *et al.*, 2008) et actualisé par quelques travaux récents, fait un inventaire des molécules de signalisation régulées par chacune des protéines SOCS. Les listes sont non exhaustives pour SOCS-1 et SOCS-3 qui, du fait de leur capacité à se fixer aux JAK, peuvent participer à la régulation d'un très grand nombre de molécules, sans spécificité de récepteur.

Tableau 2 Principales molécules de signalisation connues dans la littérature pour être inhibées par les protéines SOCS

Protéine SOCS	Molécule de signalisation régulée	Références
CIS	Érythropoïétine, GH, IL-2, IL-3, Leptine, Prolactine	(Yoshimura <i>et al.</i> , 1995 ; Emilsson <i>et al.</i> , 1999 ; Matsumoto <i>et al.</i> , 1999 ; Li <i>et al.</i> , 2000 ; Ram & Waxman, 2000 ; Endo <i>et al.</i> , 2003)
SOCS-1	IFN- $\alpha/\beta/\gamma$, IL-2, IL-4, IL-6, IL-7, IL-12, IL-15, LIF, TNF- α , etc.	(Alexander <i>et al.</i> , 1999 ; Sporri <i>et al.</i> , 2001 ; Nakagawa <i>et al.</i> , 2002 ; Cornish <i>et al.</i> , 2003 ; Chong <i>et al.</i> , 2003 ; Davey <i>et al.</i> , 2005)
SOCS-2	GH, IGF-1	(Dey <i>et al.</i> , 1998 ; Favre <i>et al.</i> , 1999)
SOCS-3	G-CSF, IL-6, LIF, IL-23, Leptine	(Lang <i>et al.</i> , 2003 ; Takahashi <i>et al.</i> , 2003 ; Croker <i>et al.</i> , 2004 ; Mori <i>et al.</i> , 2004 ; Chen <i>et al.</i> , 2006)
SOCS-4	LIF	(Sutherland <i>et al.</i> , 2012)
SOCS-5	IL-4	(Seki <i>et al.</i> , 2002)
SOCS-6	Insuline	(Krebs <i>et al.</i> , 2002)
SOCS-7	Insuline	(Banks <i>et al.</i> , 2005)

CIS : Cytokine-Inducible SH2 domain-containing protein (protéine contenant un domaine SH2 inducible aux cytokines) ; G-CSF : Granulocyte-Colony Stimulating Factor (facteur de stimulation des colonies de granulocytes) ; GH : Growth Hormone (hormone de croissance) ; IFN : InterFeroN (interféron) ; IGF : Insulin-like Growth Factor (facteur de croissance ressemblant à l'insuline) ; IL : InterLeukin (interleukine) ; LIF : Leukemia Inhibitory Factor (facteur inhibiteur de la leucémie) ; SOCS : Suppressor Of Cytokine Signaling (suppresseur de signalisation des cytokines) ; TNF : Tumor Necrosis Factor (Facteur de nécrose tumorale)

II.2.b. Cas particulier de SOCS-2

La mutation causale étudiée dans cette thèse étant située dans le domaine SH2 du gène *SOCS2* (Rupp *et al.*, 2015), nous allons à présent nous intéresser plus particulièrement aux molécules de signalisation qui peuvent être impactées par l'action de la protéine SOCS-2 (Rico-Bautista *et al.*, 2006 ; Letellier & Haan, 2016).

Comme nous venons de le voir, les protéines SOCS sont exprimées suite à la fixation d'un ligand sur son récepteur qui entraîne la cascade de phosphorylations des JAK, des STAT ainsi que la migration des STAT dans le noyau pour stimuler la transcription de gènes spécifiques (dont les gènes *SOCS*). Il a été montré que l'expression de la protéine SOCS-2 était induite *via* la voie JAK-2/STAT-5b qui peut être activée par la fixation d'un grand nombre de ligands sur leurs récepteurs, tels que la GH, l'IGF-1, la PRL (*PRoLactin* : prolactine), le LIF (*Leukemia Inhibitory Factor* : facteur inhibiteur de la leucémie), les IL (interleukines) -2, -3, et -6, l'EGF (*Epidermal Growth Factor* : facteur de croissance épidermique), etc. Bien que son expression soit induite par de nombreux ligands, il n'est pas encore tout à fait démontré que SOCS-2 soit impliquée dans leur régulation.

Moins étudiée que les protéines CIS, SOCS-1 et SOCS-3, sur lesquelles de nombreuses recherches ont été conduites en raison de la multiplicité des régulations directes, SOCS-2 est surtout connue pour son implication dans la régulation de la GH et de l'IGF-1 (Favre *et al.*, 1999) et donc dans la croissance cellulaire (Metcalf *et al.*, 2000 ; Horvat & Medrano, 2001 ; Greenhalgh *et al.*, 2002a ; Greenhalgh *et al.*, 2005 ; Dobie *et al.*, 2015). Il a été montré *in vivo* chez des souris que l'inactivation du gène *SOCS2* provoquait un dérèglement du contrôle de la GH, résultant en un phénotype de gigantisme. Les animaux chez lesquels *SOCS2* est inactif sont en effet 30 à 40% plus grands que les types sauvages à l'âge adulte, alors qu'aucune différence n'est visible à la naissance. Cette augmentation de la taille est due à un développement plus marqué et un plus grand nombre de cellules dans les organes internes et les os. Ils ressemblent phénotypiquement aux souris transgéniques GH et IGF-1. De plus, ce phénotype est inversé lorsque le gène codant pour la GH est inactif, confirmant ainsi que *SOCS2* est un régulateur négatif clé de l'augmentation de la croissance induite par la GH. A travers la régulation de la GH, *SOCS2* a également un effet sur la différenciation neuronale chez les souris (Turnley *et al.*, 2002 ; Ransome *et al.*, 2004), plus particulièrement sur le nombre et la taille de multiples types de cellules neuronales et gliales.

Le mécanisme de régulation utilisé par SOCS-2 sur cette voie n'est pas encore tout à fait élucidé, mais il semblerait qu'il s'agisse principalement du mécanisme par lequel SOCS-2 est

capable de se lier au récepteur de la GH aux mêmes sites de liaison que STAT-5, empêchant ainsi leur phosphorylation et la cascade de signalisation de l'hormone (Hansen *et al.*, 1999 ; Ram & Waxman, 1999 ; Greenhalgh *et al.*, 2002a). Cela signifie que SOCS-2 exerce un rétrocontrôle négatif sur la transduction du signal de la GH. En l'absence de SOCS-2, ce contrôle n'est plus présent et le signal transmis n'est plus régulé, d'où le phénomène de gigantisme. Plus récemment, une étude (Vesterlund *et al.*, 2011) a montré que SOCS-2 était aussi capable d'inhiber la GH *via* la dégradation par le protéasome du récepteur de la GH complexé à SOCS-2, grâce au recrutement du complexe d'ubiquitines (Bullock *et al.*, 2006 ; Kung *et al.*, 2019). Bien qu'à faible dose, SOCS-2 ait un effet inhibiteur sur l'action de la GH, il a été montré qu'à forte dose, SOCS-2 pouvait au contraire intensifier le signal de l'hormone (Favre *et al.*, 1999 ; Greenhalgh *et al.*, 2002b). L'explication sur cette dualité d'effets reste largement non expliquée mais il semblerait que cet effet soit lié à sa capacité d'inhiber l'action d'autres protéines SOCS régulant la GH en les dirigeant vers la voie de dégradation du protéasome.

La capacité de SOCS-2 à se lier à certains récepteurs est assez bien décrite, comme c'est le cas pour le récepteur de la GH sur le site de phosphorylation pY595 (Hansen *et al.*, 1999), de l'IGF-1 (Dey *et al.*, 1998), de l'érythropoïétine sur le site pY426 (Eyckerman *et al.*, 2001), de la leptine sur le site pY1077 (Lavens *et al.*, 2006), et de l'EGF (Goldshmit *et al.*, 2004). Cependant, bien que SOCS-2 puisse se lier aux récepteurs de ces ligands, il n'est pas encore tout à fait clair à ce jour si SOCS-2 est impliquée dans leur régulation.

SOCS-2 serait impliquée dans la régulation de certaines molécules de signalisation *via* son interaction avec les autres protéines SOCS. Par exemple, SOCS-2 régulerait la PRL grâce au blocage de l'action inhibitrice de SOCS-1 (Pezet *et al.*, 1999), la PRL étant une hormone jouant un rôle dans la lactation et la reproduction. Elle régulerait également les fonctions de l'IL-2 et de l'IL-3, impliquées dans la réponse immunitaire, *via* la dégradation de SOCS-3 (Tannahill *et al.*, 2005). Des effets similaires se retrouvent également sur la signalisation des interférons, également impliqués dans les fonctions immunitaires, et de la leptine (Piessevaux *et al.*, 2006), impliquée dans la fonction digestive.

D'autres régulations directes de SOCS-2 ont été mentionnées précédemment telles que la régulation du facteur de croissance épidermique EGF (Goldshmit *et al.*, 2004), de l'insuline (Lebrun *et al.*, 2010), hormone impliquée dans la régulation de la glycémie dans le sang, de la PRL (Harris *et al.*, 2006), ou encore celle des IL-1 β et IL-10 (Posselt *et al.*, 2011), impliquées

dans la réponse immunitaire. Cependant, les mécanismes de régulation restent encore très mal connus et d'autres études sont nécessaires pour confirmer ceux déjà identifiés.

II.3. Les fonctions biologiques régulées par SOCS-2

La protéine SOCS-2 est impliquée dans la régulation de nombreuses molécules de signalisation qui touchent différentes fonctions biologiques. Nous allons à présent nous intéresser aux études qui établissent un lien direct entre la protéine SOCS-2 et la régulation de fonctions biologiques d'intérêt pour l'élevage : la reproduction, la croissance et la production, ainsi que la santé.

II.3.a. Reproduction

La voie JAK/STAT, et par extension, les protéines SOCS semblent être impliquées dans la régulation de la gestation des mammifères en raison des molécules de signalisation sécrétées par l'embryon et par l'utérus au cours de la gestation (Fitzgerald *et al.*, 2009) telles que :

- l'IFN- γ (*InterFeroN-gamma*: interféron-gamma) et l'IL-4, régulés par SOCS-1, et impliqués dans le maintien de la gestation *via* leur influence sur la différenciation des lymphocytes T auxiliaires (Fujimoto *et al.*, 2002),
- le LIF, impliqué dans la différenciation du trophoblaste, et régulé par SOCS-3 (Boyle & Robb, 2008),
- l'IFN- τ (*InterFeroN-tau*: interféron-tau), impliqué dans la signalisation de la gestation chez les ruminants, régulé par CIS, SOCS-1 à -3 (Sandra *et al.*, 2005).

A notre connaissance, seul l'IFN- τ a été associé à une régulation impliquant SOCS-2. En effet, chez les ovins, l'IFN- τ est sécrété par l'embryon en début de gestation, entre 10 et 22 jours après l'accouplement, afin de signaler sa présence dans l'utérus maternel et permettre l'implantation (Martal *et al.*, 1998). Le signal de l'IFN- τ transitant *via* la voie JAK/STAT d'une manière similaire aux IFN- α / β / γ , une étude d'expression a montré que les protéines SOCS étaient impliquées dans la régulation de l'implantation embryonnaire (Sandra *et al.*, 2005). SOCS-1 et -2 en particulier seraient impliquées dans la régulation de la réceptivité de l'endomètre (paroi interne de l'utérus) et de la fixation de l'embryon (au stade blastocyste), tandis que CIS et SOCS-3 participeraient à la morphogenèse de la glande endométriale, à l'activité sécrétoire utérine et à la différenciation placentaire.

Une étude similaire réalisée dans l'espèce bovine a également démontré l'induction de SOCS-2 par l'IFN- τ et sa surexpression au 20^{ième} jour de la gestation, qui correspond à la date de l'implantation de l'embryon (Carvalho *et al.*, 2014). Cette étude semble indiquer un rôle

régulateur de SOCS-2 dans la gestation mais le mécanisme n'est pas décrit et reste encore à élucider.

II.3.b. Croissance et production

La protéine SOCS-2 étant bien caractérisée comme régulatrice principale de la GH, de nombreuses études se sont intéressées aux effets de SOCS-2 sur la croissance (voir paragraphe sur le gigantisme de la partie II.2.b). Chez l'Homme, un SNP localisé dans le gène *SOCS2* a été associé à la taille au moment de la puberté et à l'âge adulte (Sovio *et al.*, 2009), suggérant le rôle de SOCS-2 sur la croissance.

Chez les ovins, la mutation R96C (Rupp *et al.*, 2015) annihile la fixation de *SOCS2* au motif contenant la phosphotyrosine du récepteur de la GH. Les animaux porteurs de cette mutation ont un poids supérieur de 18%, à trois ans d'âge, en comparaison aux animaux de génotype sauvage. La taille des animaux porteurs de la mutation était augmentée de 24%. Cette même étude rapporte également une augmentation de la production laitière (+4,4% chez les animaux porteurs). Deux hypothèses sont émises concernant cette augmentation de la production : soit elle est associée à l'augmentation du format, et donc proportionnellement à la capacité productive de la glande mammaire sous l'impulsion de la GH (Zhang *et al.*, 2014), soit elle serait directement un résultat de la dérégulation de la PRL et de l'activité de STAT-5, qui jouent un rôle primordial dans le développement de la glande mammaire (Harris *et al.*, 2006). Plus récemment, une étude a affirmé une tendance à l'augmentation de la croissance d'animaux génétiquement modifiés pour cette même mutation (Zhou *et al.*, 2019). Cette étude repose cependant sur un nombre très restreint d'animaux (3 animaux modifiés et 5 animaux témoins) à un jeune âge (naissance, 30 jours et 60 jours d'âge), la différence étant plus marquée à 60 jours, laissant supposer que la différence pourrait être accrue à l'âge adulte.

Par ailleurs, le gène *SOCS2* a été détecté comme gène candidat sous sélection dans un panel de 27 races ovines d'Europe, avec un haplotype présent principalement dans les races du nord de l'Europe (Rochus *et al.*, 2018), suggérant une sélection qui aurait été réalisée au cours de l'histoire évolutive des races pour un phénotype d'intérêt pour les éleveurs, comme la production, le format, ou encore la reproduction.

II.3.c. Santé et réponse immunitaire

Chez l'Homme, une faible expression du gène *SOCS2* a été associée à de nombreux cancers tels que le cancer du sein, des poumons, du foie et de l'ovaire (Wikman *et al.*, 2002 ; Sutherland *et al.*, 2004 ; Farabegoli *et al.*, 2005 ; Haffner *et al.*, 2007 ; Qiu *et al.*, 2013 ; Zhu *et al.*, 2013).

SOCS-2 a également été identifiée comme fortement impliquée dans la différenciation des cellules dendritiques (Hu *et al.*, 2009 ; Posselt *et al.*, 2011), dans l'expansion des lymphocytes T auxiliaires de type 2 (T *helper* 2, ou Th2) et le développement de la réponse allergique de type 2 (Knosp *et al.*, 2011), et dans la différenciation des cellules NK (Kim *et al.*, 2017), toutes ces cellules étant des effecteurs majeurs des réponses immunitaires innée et adaptative. Par ailleurs, la dernière étude mentionnée (Kim *et al.*, 2017) est la première étude à notre connaissance à démontrer la capacité de SOCS-2 à se lier à une protéine JAK (JAK-2) dans son mécanisme de régulation des cellules NK.

Chez des souris déficientes pour SOCS-2, la production de cytokines pro-inflammatoires est incontrôlée (Machado *et al.*, 2006). L'infection expérimentale de souris par le virus BoHV-5 (*Bovine herpesvirus 5* : herpès-virus bovin 5) induit une méningo-encéphalite dans laquelle intervient SOCS-2 (Aparecida Silva Barbosa *et al.*, 2016). Les résultats ont montré que les souris déficientes pour SOCS-2 ont une inflammation cérébrale exacerbée et présentent une expression réduite de l'ARN messager cérébral de l'IFN- α et de l'IFN- β ainsi qu'une expression accrue de l'ARN messager de SOCS-1 par rapport aux souris de type sauvage.

La mutation ovine *SOCS2* R96C a un effet très fort sur l'inflammation mammaire (CCS), comme décrit précédemment (Rupp *et al.*, 2015). *SOCS2* a également été cité parmi les gènes candidats associés à des phénotypes de résistance aux parasites gastro-intestinaux chez les ovins (Gossner *et al.*, 2013 ; Benavides *et al.*, 2016). La voie JAK/STAT a aussi été détectée dans des analyses d'association chez les bovins en lien avec le caractère SCS (Usman *et al.*, 2014).

III. Les études génétiques et génomiques sur les mammites ovines

III.1. Objectifs et démarches

Cette partie dresse un état des lieux complet des connaissances sur la génétique et la génomique des mammites chez les ovins, afin de replacer l'importance de la mutation *SOCS2* et des autres QTL dans le déterminisme du caractère dans la race de découverte - Lacaune - et dans les autres races ovines laitières. Elle met également en lumière les relations génétiques de ce caractère complexe de résistance aux mammites avec les différents caractères d'intérêt en ovin.

Cette partie est rédigée sous la forme d'un article de synthèse bibliographique accepté pour publication le 19 mai 2019 dans un numéro spécial de la revue *Small Ruminant Research* intitulé *Production of quality milk-Mastitis* (Production de lait de qualité - mammites). Les tableaux additionnels (Table S1) de cet article sont présents en annexe (Annexe 1).

III.2. Résumé de l'article de synthèse bibliographique

III.2.a. Synthèse

Les mammites sont une préoccupation financière et de bien-être animal en ovins laitier et allaitant. Dans cet article de synthèse, nous résumons les progrès récents de la recherche sur la génomique de la résistance aux mammites en ovin. Les héritabilités estimées pour les caractères liés aux mammites tels que le SCS ou le comptage bactériologique du lait ont confirmé une base génétique pour la résistance aux mammites. Les résultats récents d'études génomiques (études d'associations pangénomiques, identification de mutations causales, études sur l'expression des gènes) ont mis en évidence des régions du génome et des gènes et mécanismes possibles sous-jacents à ce caractère de résistance. Une partie des régions génomiques était commune parmi les races et les populations, et témoigne du partage partiel des mécanismes génétiques liés aux mammites entre différentes populations éloignées d'ovins laitiers. L'accumulation de données génétiques a toutefois mis en évidence la nature polygénique du caractère et la complexité du phénotype de résistance. Les analyses quantitatives et génomiques ont révélé des compromis et des synergies sous contrôle génétique entre la résistance aux mammites et d'autres facteurs d'efficacité (production laitière) et de résilience (phénotypes de morphologie mammaire et de maladie métabolique). Nous avons également signalé comment les informations phénotypiques, généalogiques et génomiques ont été utilisées dans les schémas de sélection ovins, l'objectif étant d'améliorer la santé et le bien-être des animaux, la qualité hygiénique des produits laitiers et l'efficacité et la résistance.

III.2.b. Introduction

Les études génétiques sur le caractère de résistance aux mammites ont débuté dans les années 1990 principalement sur les ovins laitiers. Ces études portent sur des estimations de l'héritabilité d'un caractère reflétant l'infection mammaire, le SCS. Puis, grâce à l'avancée des outils génomiques (puces à SNP), des études d'association et d'expression de gènes ont été réalisées, aboutissant à une application de ces résultats dans les schémas de sélection.

III.2.c. Les paramètres génétiques des mammites ovines

Des paramètres génétiques sur différents caractères liés aux mammites ont été estimés : SCS (majoritairement), CMT (*California Mastitis Test* ou test au Teepol), occurrence des mammites cliniques (caractère binaire), analyses bactériologiques du lait. Les héritabilités sont faibles et varient entre 0.02 et 0.28. Ces différents caractères sont bien corrélés positivement entre eux. Les corrélations génétiques de ces caractères liés aux mammites avec :

- les caractères de production laitière sont mitigés dans la littérature (positives ou négatives selon les races)
- les caractères de morphologie mammaire et de facilité à la traite montrent que des brebis avec un pis pendant et avec des trayons plus longs et/ou capables de libérer le lait plus rapidement risquent davantage de développer une mammite
- les caractères de conformation sont proches de 0, indiquant qu'il n'y aurait pas de lien génétique
- les autres caractères de résistance aux maladies sont mitigés : pas de corrélation avec la résistance aux nématodes gastro-intestinaux, un lien faible positif avec le métabolisme énergétique

III.2.d. Les bases de la génomique pour la résistance aux mammites

Les découvertes de QTL sont réalisées principalement sur le caractère de SCS à l'aide de marqueurs (anciennement microsatellites et plus récemment SNP). Cet article fait un inventaire de toutes les régions QTL détectées dans trois populations laitières. Suite à ces premières détections, quelques études ont poursuivi les recherches en réalisant un séquençage fin afin de découvrir des gènes candidats et/ou des mutations causales associé(e)s, comme c'est le cas par exemple de la mutation causale dans le gène *SOCS2* sur le chromosome 3. Ces QTL ont ensuite été validés par plusieurs études dans des populations différentes et ceci a permis d'affirmer le partage partiel des mécanismes génétiques liés aux mammites. Les mécanismes sous-jacents ont également été étudiés grâce à des données de gènes différentiellement exprimés après challenge ou infection expérimentale.

III.2.e. L'élevage pour l'amélioration de la résistance aux mammites

Les schémas de sélection chez les ovins laitiers ont commencé à enregistrer à grande échelle le caractère SCS à partir des années 1980. A ce jour, ce caractère est intégré dans les objectifs de sélection dans une seule race en France (25% de l'objectif en complément de caractères de production laitière et de morphologie mammaire), dans un but d'amélioration de la santé mammaire, mis en place suite à une étude sur des lignées divergentes pour le SCS qui montrait la réussite d'une telle sélection. Ce caractère de résistance aux mammites est aussi de plus en plus à l'étude chez les ovins allaitants où les enregistrements sont plus difficiles (pas de prélèvement de lait journalier) et la sélection pourrait se tourner vers le caractère CMT plus facile à collecter. De manière pratique, il n'est pas possible d'utiliser l'information directe de gènes majeurs pour cette sélection car le caractère de résistance aux mammites est complexe et polygénique. En revanche, l'apport de la génomique dans les modèles d'évaluation amenant des précisions augmentées pour les valeurs génétiques prédites est encourageant. Des méthodes existent pour insérer l'information des QTL connus dans les modèles d'évaluation et pourraient être appliquées pour améliorer la santé mammaire des ovins.

III.3. Article de synthèse bibliographique (Oget *et al.*, 2019c)



Genetic and genomic studies in ovine mastitis

C. Oget^{*}, G. Tosser-Klopp, R. Rupp

GenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France

ARTICLE INFO

Keywords:

Genetic resistance
Intramammary infections
Mastitis
Animal genomics
Sheep
Genetic parameters

ABSTRACT

Mastitis is a strong financial and animal welfare concern in both dairy and meat-producing sheep. In this review article, we summarized recent advances in research on genomic of mastitis resistance in sheep. Heritability estimates for mastitis-related traits such as somatic cell scores or milk bacteriological counting have confirmed a genetic basis for mastitis resistance. Recent outputs from genomic-based studies comprising genome-wide associations, the identification of one causal mutation as well as gene expression studies have highlighted regions of the genome, and possible genes and mechanisms underlying the resistance trait. Part of genomic regions was common among breeds and populations, and have testified to the partial sharing of mastitis-related genetic mechanisms between different distant dairy sheep populations. Accumulating genetic data, however, have underlined the polygenic nature of the trait and the complexity of the resistance phenotype. Both quantitative and genomic analyses have further revealed trade-offs and synergies under genetic control between mastitis resistance and other efficiency (milk production) and resilience related traits (udder type traits and metabolic disease). We have also reported how phenotypic, pedigree and genomic information has been used in sheep breeding, the aim being to improve the animals' health and welfare, the hygienic quality of milk products and overall efficiency and resilience.

1. Introduction

Genetics of mastitis has been studied more recently in dairy sheep (Baro et al., 1994; El-Saied et al., 1999; Barillet et al., 2001; Serrano et al., 2003) than in dairy cattle where it has been increasingly well documented since the 1980s (Emanuelson and Philipsson, 1984; Emanuelson et al., 1988; Mrode and Swanson, 1996; Heringstad et al., 2000; Detilleux, 2002; Rupp and Boichard, 2003). In general, relevant studies have focused on dairy sheep, with few data produced in meat or wool production sheep populations.

First evidence that a host's response to mastitis was under genetic control, came from quantitative genetic studies and the estimation of genetic parameters for mastitis-related traits. The approach relies on the assumption that the trait is determined by a combination of animal (age, lactation stage, etc.), genetic (breed, line, etc.), and environmental factors, the latter including both microbiological (infection pressure, pathogen species and strain, etc.) and husbandry parameters (breeder practices regarding machine milking, sheepfold, etc.). No assumption had been made about the underlying genes and mechanisms. The heritability parameter quantifies the amount of variability in a given trait that is due to additive genetic values. It also allows the

estimation of genetic value for the animals (breeding value) and predicting to what extent a selection process can be efficient. These studies, however, require large data sets with phenotypic information and pedigree. Therefore, the most common mastitis-related trait used in ovine mastitis genetic studies is the somatic cell count (SCC). Indeed, somatic cell count can be easily measured in milk samples, as an indirect predictor of mammary infection, which is much more frequent than clinical mastitis in sheep (Bergonier et al., 2003).

Since 2009, the development of genomic methods and tools applied to sheep have contributed to enhance insights into the genetic control of processes governing host resistance to mastitis. The sheep genome was sequenced in 2007 (Jiang et al., 2014) and several low- to high-density ovine SNP (single-nucleotide polymorphism) chips have been developed as part of the International Sheep Genomics Consortium (ISGC; www.sheepmap.org) (Kijas et al., 2009). Availability of such tools has enabled genome scans to identify and fine map genomic regions associated with traits (QTL for Quantitative Trait Loci), the gene expression studies to highlight pathways and genes involved in the host's response to infection, and the first use of genome-wide information in sheep breeding.

This paper reviews the state of the art on the genetic control of

^{*} Corresponding author at: INRA - GenPhySE-Génétique, Physiologie et Systèmes d'Élevage UMR1388, 24 Chemin de Borde Rouge – Auzeville CS 52627, 31326 Castanet Tolosan cedex, France.

E-mail address: claire.oget@inra.fr (C. Oget).

<https://doi.org/10.1016/j.smallrumres.2019.05.011>

Received 18 March 2019; Received in revised form 17 May 2019; Accepted 19 May 2019

Available online 23 May 2019

0921-4488/© 2019 Institut National de la Recherche Agronomique. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
Estimates of heritabilities for mastitis-related traits and genetic correlations with milk yield and mastitis-related traits in ewes.

Reference (Chronological order)	Sheep breed	Number of		Trait	Model	Heritability	Genetic correlations: trait and	
		Records	Ewes				Milk yield	Other mastitis traits
Baro et al., 1994	Churra	10,171 test-days	3,832	Somatic cell scores	Test-day	0.04	-0.37	-
El-Saied et al., 1999	Churra	3,231 lactation periods	2,379	Somatic cell scores	Lactation period	0.12	-0.15	-
Barillet et al., 2001	Lacaune	23,091 test-days	5,272	Somatic cell scores (1 st lactation period)	Lactation period	0.15	0.11	-
Othmane et al., 2002a	Churra	1,962 lactation periods	1,111	Somatic cell scores	Lactation period	0.11	-0.17	-
Othmane et al., 2002b	Churra	7,492 test-days	1,119	Somatic cell scores	Test-day	0.11	-0.36	-
Gonzalo et al., 2003	Churra	4,692 test-days	1,337	Somatic cell scores	Test-day	0.09 - 0.11	-	-
Flupp et al., 2003	Lacaune	273,975 test-days	94,445	Somatic cell scores (1 st lactation period)	Lactation period	0.13	0.18	SCS (2 nd lactation period): 0.93
Serrano et al., 2003	Manchega	131,947 test-days	45,499	Somatic cell scores (2 nd lactation period)	Lactation period	0.12	0.08	SCS (1 st lactation period): 0.93
		37,641 test-days	10,962	Somatic cell scores (1 st lactation period)	Lactation period	0.12	-0.12	-
		36,793 test-days	10,560	Somatic cell scores (2 nd lactation period)	Lactation period	0.19	-0.14	-
		29,358 test-days	8,379	Somatic cell scores (3 rd lactation period)	Lactation period	0.24	-0.15	-
Hamann et al., 2004	East Friesian	9,729 test-days	1,108	Somatic cell scores	Test-day	0.16	0.02	-
Legarra and Ugarte, 2005	Latxa	9,805 lactation periods	6,165	Somatic cell scores	Lactation period	0.13	-0.30	-
Barillet, 2007	Lacaune	121,283 lactation periods	121,283	Somatic cell scores (1 st lactation period)	Lactation period	0.15	0.15	-
		13,066 test-days	2,277	Somatic cell scores (1 st lactation period)	Test-day	0.14	0.23	-
		163,458 test-days	58,378	Somatic cell scores (1 st lactation period)	Lactation period	0.10	0.21	-
Caseo et al., 2010	Sarda, Sarda × Lacaune	1,682 test-days	1,587	Somatic cell scores (1 st lactation period)	Lactation period	0.26 - 0.28	-	-
Riggio et al., 2010	Valle del Belice	8,843 test-days	1,120	Somatic cell scores	Test-day	0.09	-	-
				Infection status (bacteriological examination)	Threshold animal	0.09	-	SCS infected / uninfected ewes: 0.81 / 0.51
		2,866 test-days	805	Somatic cell scores of infected ewes	Test-day	0.03	-	InfS / SCS uninfected ewes: 0.81, 0.62
		5,977 test-days	1,062	Somatic cell scores of uninfected ewes	Test-day	0.10	-	InfS / SCS infected ewes: 0.51 / 0.62
De la Fuente et al., 2011	Churra	10,189 test-days	3,977	Somatic cell scores	Test-day	0.09	-	InfS: 0.93
Tolone et al., 2013	Valle del Belice	17,843 test-days	2,040	Somatic cell scores	Test-day	0.09	-	SCS: 0.93
				Infection status (bacteriological examination)	Threshold animal	0.09	0.59	InfS by cns / by Str.: 0.92 / 0.36
Tolone et al., 2016	Valle del Belice	5,305 lactation periods	2,350	Infection status (bacteriological examination)	Threshold animal	0.02	-	InfS / InfS by Str: 0.92 / 0.24
				Infection status by coagulase-negative staphylococci (bacteriological examination)		0.02	-	InfS / InfS by cns: 0.36 / 0.24
				Infection status by <i>Streptococcus</i> spp. (bacteriological examination)		0.09	-	InfS / InfS by cns: 0.36 / 0.24
Baños et al., 2017	Chios	2,436 samples	609	Somatic cell scores (1 st and 2 nd lactation period)	Test-day (week)	0.11	-0.12	CMT / TBC / Dev c/m: 0.77 / 0.56 / 0.35
				Scores of California Mastitis Test (1 st and 2 nd lactation periods)		0.12	-0.12	CMT / TBC / Dev c/m: 0.77 / 0.21 / 0.36
				Total bacterial counts (1 st and 2 nd lactation periods)		0.09	-0.11	SCS / CMT / Dev c/m: 0.56 / 0.21 / 0.35
				Development of clinical mastitis (1 st and 2 nd lactation periods)	Logit function	0.18	-0.09	SCS / CMT / TBC: 0.35 / 0.36 / 0.35
O'Brien et al., 2017	Belclare, Charollais, Suffolk, Texel, Vendéen	3,378 examinations	3,378	Scores of California Mastitis Test	Linear animal	0.04	-	-
Alluin et al., 2018	Lacaune	-	377,945	Somatic cell scores (1 st lactation period)	Lactation period	0.22	0.12	StaphC: 0.72
			518	Staphylococcal counts (1 st lactation period)	Lactation period	0.18	0.31	SCS: 0.72

(continued on next page)

Table 1 (continued)

Reference (Chronological order)	Sheep breed	Number of Records	Trait	Model	Heritability	Genetic correlations: trait and		
						Milk yield	Other mastitis traits	
McLaren et al., 2018	Texel	3,410 samples 3,539 samples 3,529 samples	2,957 Ewes	Average somatic cell scores of samples from both mammary glands Sum of scores of California Mastitis Test in samples from both mammary glands Maximum score of California Mastitis Test in samples from both mammary glands	Animal model	0.08 - 0.11 0.09 - 0.11 0.07 - 0.08	- - -	Sum CMT: 0.76 - 0.96 Max CMT: 0.79 - 0.98 Av SCS: 0.76 - 0.96 Max CMT: 0.99 Av SCS: 0.79 - 0.98 Sum CMT: 0.99

SCS: Somatic Cell Score, InfS: Infection Status, cns: coagulase-negative *Staphylococci*, Str.: *Streptococci*, CMT: scores of California Mastitis Test, TBC: Total Bacterial Counts, Dev c/m: Development of clinical mastitis, StaphC: *Staphylococcal* Counts, Sum CMT: Sum of scores of California Mastitis Test in samples from both mammary glands, Max CMT: Maximum score of California Mastitis Test in samples from both mammary glands, Av SCS: Average Somatic Cell Scores of samples from both mammary glands.

* Results obtained by Banos et al. (pre-print on the BioRxiv platform, doi: <https://doi.org/10.1101/577015>) that used the same samples as in Banos et al. (2017).

resistance to mastitis, with primary reference to sheep. Genetic parameters for different mastitis-related traits are presented and genetic associations with other traits of interest are discussed. The paper then reports on recent outputs from genomic-based studies comprising QTL detection studies and the identification of causal mutations as well as gene expression studies. Finally, it considers how information using genome-wide information (or not) is currently used in sheep breeding with some discussion on future prospects.

2. Genetic parameters of ovine mastitis

2.1. Genetic parameters for mastitis-related traits

For three decades, published literature on genetic parameters of ovine mastitis in dairy sheep has been accumulating, as reported in earlier reviews (Carta et al., 2009; Rupp and Foucras, 2010; Riggio and Portolano, 2015). The most common mastitis-related traits used in genetic studies is a logarithmic transformation of somatic cell count termed 'Somatic Cell Score' (SCS) (Ali and Shook, 1980). Heritabilities for somatic cell scores (lactation period average or repeated test day records) have been reported in sheep, as shown in Table 1. Heritability estimates are low and range from 0.04 to 0.28 with a weighted average of 0.163 (Table S1 for the explanation of the calculation) for various dairy sheep breeds, including Chios (Greece), Churra (Spain), East Friesian (Germany), Lacaune (France), Latxa (Spain), Manchega (Spain), Manech Red Face (France), Sarda (Italy) and Valle del Belice (Italy). These heritabilities showed that genetic variation existed for this trait, although to a lesser extent than for production traits with heritabilities ranging from 0.30 to 0.60. The California Mastitis Test score is another less used indirect mastitis-related trait based on milk somatic cells. This test consists of the addition in the milk of a reagent composed of a detergent and a pH indicator. When mixed with the milk, the reagent reacts and forms a viscous gel, which can be scored using a range of milk color and viscosity, thereby indicating the number of somatic cells and the level of inflammation. Only one study in dairy sheep (Banos et al., 2017) estimated a heritability for this trait (0.12), which is in the same range as for somatic cell score. However, both somatic cell scores and results of the California Mastitis Test are two indirect measures for mastitis, which reflect the degree of inflammation in the mammary gland rather than infection status. The trait, therefore, combines the measure of severity of infection and intensity of the host's response. Direct measures of mastitis, e.g., the occurrence of clinical mastitis based on examinations of veterinarians have been very little studied in dairy sheep, because of the small frequency of the disease (Bergonier et al., 2003) and the difficulty to implement it on a large scale. Heritability of this binary trait (presence/absence of clinical mastitis) has been estimated at 0.18 in the Chios breed (Banos et al., 2017), which is also in the same range as for somatic cell scores.

Very recently, studies in meat-producing sheep breeds have estimated genetic parameters for mastitis-related traits, as it is a strong financial and animal welfare concern (Conington et al., 2008). Heritability of occurrence of clinical mastitis was estimated at 0.04 in an Irish population including Belclare, Charollais, Suffolk, Texel and Vendean breeds (O'Brien et al., 2017). In the Texel breed, somatic cell scores and results of California Mastitis Test estimates of heritability ranged from 0.07 to 0.11 (McLaren et al., 2018). Although these heritabilities are somewhat smaller than in dairy sheep, probably due, at least partly, to the recording scheme (only one examination per animal), they have confirmed that, as expected, genetic control for mastitis also exists in those meat-producing sheep breeds.

During the last decade, genetic studies have focused on bacteriological analyses of milk samples, in order to better quantify the intramammary bacterial pressure in dairy sheep. The most widely used technique in the dairy sheep literature is a count of bacterial colonies in milk samples after incubation on agar plates. This conventional technique leads to a binary trait of infection status (ewe considered as

infected if over 5 colony-forming-units per 10 μ L of milk of one bacterial species would be recovered). Estimates of heritability of this trait termed 'infection status' ranged from 0.02 to 0.09 (Table 1) in the Valle del Belice breed (Riggio et al., 2010; Tolone et al., 2013, 2016) and seem to depend on the species of bacteria. One study (Banos et al., 2017) has estimated heritability of another trait based on bacteriological examination of milk samples, the total bacterial count, which refers to the number of viable bacteria per mL of milk. Heritability of this estimate is of the same order of magnitude as the infection trait (0.09), which is lower than heritabilities of the traits based on milk somatic cells. One last technique, recently employed in the French Lacaune breed (Allain et al., 2018; Oget et al., 2019) has been based on amplification on staphylococcal DNA by qPCR in milk samples, representing the number of bacterial genomes in 10 μ L of milk. Estimate of heritability of this continuous trait was in the same range as somatic cell scores (0.18).

Genetic correlations have also been estimated (Table 1) between the various mastitis-related traits described above. The occurrence of clinical mastitis has been found to be positively correlated with somatic cell scores, results of California Mastitis Test and total bacterial counts (0.35–0.36) in Chios animals (Banos et al., 2017), suggesting a common genetic basis between development of disease and the various traits indirectly indicating mastitis. Genetic correlations between the two leucocyte-related traits (results of California Mastitis Test and somatic cell scores) ranged from 0.76 to 0.98 (Banos et al., 2017; McLaren et al., 2018), confirming the usefulness of the California Mastitis Test as an alternative to cell counting. Infectious status (conventional technique) and somatic cell scores have also been found to be highly correlated (0.51 to 0.93) (Riggio et al., 2010; Tolone et al., 2013), as did also total bacterial counts with somatic cell scores (0.56) (Banos et al., 2017) and staphylococcal DNA with somatic cell scores (0.72) (Allain et al., 2018). Milk bacteriological traits seemed then to be genetically close to the most used mastitis-related trait 'somatic cell score'. These reasonably high values have suggested a commonality of resistance mechanisms that lead to better resistance to either persistent intra-mammary infection (continuously high SCC) or acute clinical episodes with presence of pathogens in the udder, despite the fact that those forms of mastitis might be associated with various environmental conditions, pathogens, and status of the animal.

2.2. Genetic relationships of mastitis resistance with other traits

2.2.1. Relationships with milk production traits

Genetic correlation estimates between mastitis-related traits and milk yield trait in sheep are presented in Table 1. These are inconsistent across the various dairy breeds. Indeed, estimates were found to be negative (i.e., favorable) for Spanish and Greek breeds, with genetic correlations from -0.37 for Churra to -0.09 for Chios ewes, and positive (i.e., unfavourable) for German, Italian and French breeds, with genetic correlations 0.02 for East Friesian to 0.59 for Valle del Belice ewes. The weighted average of these estimates is 0.078 (Table S1). An explanation for these inconsistent results between the different breeds may simply be statistical reasons. Differences in genetic correlation estimates between Spanish and the other breeds can be due to lower selection pressure applied in the Spanish breeds, or to the fact that estimates in Spanish breeds were mainly obtained considering different lactations whereas, in the other breeds, they mostly considered only the first lactation. Some pleiotropic genes, for example the group-specific component (GC) gene, could explain the relation between udder health and production traits. This gene encodes the vitamin D-Binding Protein (DBP) which has multiple roles in immune defense and milk production, as found in cattle (Olsen et al., 2016). Another pleiotropic gene is the *suppressor of cytokine signalling2* (SOCS2) gene, which has been found to be associated with mastitis in sheep and also had an effect on growth and milk production (Rupp et al., 2015). It also can be due to a selection sweep, when selection for milk production has 'hitchhiked'

along with it a nearby allele on the chromosome that is associated with mastitis susceptibility. Other possible explanations for strong unfavorable genetic correlations are the biological competition between functions for energy and nutrients and the relationship between mastitis and udder type traits.

2.2.2. Relationships with udder type traits and milking ease

A few studies in dairy and meat-producing sheep have dealt with estimating genetic correlations between udder-type and mastitis-related traits (Legarra and Ugarte, 2005; Barillet, 2007; Casu et al., 2010; McLaren et al., 2018). Udder attachment was negatively correlated with mastitis (genetic correlations: -0.27 to -0.42), whereas long teats were favorable for mastitis (genetic correlations: 0.26 to 0.44), indicating that ewes with a pendulous udder or longer teats are at greater risk to develop mastitis. Moreover, teat angle has also been negatively correlated with mastitis (genetic correlations: -0.12 to -0.55). One study also has demonstrated the positive genetic relationship between mastitis resistance and milking ease (Allain et al., 2018). Indeed, genetic correlations of somatic cell scores or quantity of staphylococcal DNA in milk with milk flow traits ranged from 0.26 to 0.30 and from -0.04 to -0.33 with latency time and real milking time, respectively. These values confirmed that ewes with the higher relaxing ability of the teat muscle (able to quickly release milk) allowed more pathogens to enter into the mammary gland, potentially leading to mastitis.

2.2.3. Relationships with body, feet type and meat production traits

Genetic association of mastitis traits with body conformation, feet type, and meat production traits have been found to be close to zero. Indeed, the estimate of genetic correlation of somatic cell scores with litter size was at 0.001 for East Friesian ewes (Hamann et al., 2004). Estimates with morphological traits, including stature, rear legs, feet angle, rump width, and general body score, ranged from -0.06 to 0.11, with high standard errors for Churra ewes (De la Fuente et al., 2011), as well as for meat-producing sheep breeds (O'Brien et al., 2017). These results indicate that selection for one of these traits would not affect susceptibility to mastitis of the individuals.

2.2.4. Relationship with resistance to other diseases

Mastitis is mainly caused by extracellular and facultative intracellular bacteria and the host's resistance probably involves a crucial role of the type-1 immune response. However, the control of parasites is predominately governed by a type-2 immune response. Also, there is evidence that different components of the immune response, including type-1 and type-2 responses, and resistance to various diseases, show null to antagonistic genetic correlation (Mouton et al., 1984; Pinaud van der Laan, 2002; Crawley et al., 2005). In sheep, a link between mastitis resistance with resistance to gastrointestinal nematodes was reported in an experiment, in which sheep divergently selected for either high or low somatic cell counts received an oral challenge with *Haemonchus contortus* (Traoré et al., 2008). Whereas lambs showed highly variable response to challenge, there was no difference between the genetic lines, which suggested that genetic resistance to either mastitis or gastrointestinal nematodes were under independent genetic regulations.

The link of mastitis resistance to metabolic disease was addressed in an experiment, in which sheep divergently selected for either high or low somatic cell counts, received an energy-restricted diet (Bouvier-Muller et al., 2016, 2018). The authors showed that sheep from the low cell count line, which were less susceptible to mastitis, were also less susceptible to ketosis in early lactation (identified by measurements of concentrations of β -hydroxybutyrate and non-esterified fatty acids in blood). They concluded that these positive correlations suggested some commonalities in genetic control of immune response and energy metabolism, but may also reflect indirect associations due to competition for nutrients.

Finally, a question was put forward whether resistance to mastitis

was independent to resistance to other microbial diseases of sheep, e.g., footrot, paratuberculosis or *Small Ruminant Lentivirus* infection (Davies et al., 2009), but has not been addressed properly, due to lack of large-scale recording for those diseases.

3. Genomic basis for mastitis resistance

3.1. QTL detection

3.1.1. Primary discovery of QTLs

QTL for mastitis resistance have been evidenced in dairy sheep. Among primary QTL detection implemented within the 'Genesheepsafety' EU project, the common mastitis resistance phenotype was milk somatic cell count, measured periodically over a lactation period. Studies used low-density microsatellite panel (130 - 181 markers) at the time and only gave a limited resolution of the QTL locations and large chromosomal regions were not covered. One major finding was a QTL on *Ovis aries* chromosome 20 (OAR20), 30 cM distant from the MHC (Major Histocompatibility Complex) gene cluster, in a Churra population with 1421 ewes (Gutiérrez-Gil et al., 2007).

The availability of the Illumina OvineSNP50 beadchip, with 54 K SNPs in 2009 has allowed undertaking genome-wide association studies to map more precisely *loci* describing significant genetic variations in mastitis susceptibility. The 'Sustainable Solutions for Small Ruminants' ('3SR') project has been initiated, with the aim to implement genome scans for traits with high importance for health and sustainability in sheep, including mastitis resistance, and to develop selectable genetic markers and identify the causative mutation(s) where possible. QTL detection for somatic cell count was performed in three large dairy sheep populations by the consortium in France, Spain, and Italy (Sechi et al., 2013; Rupp et al., 2015; Gutiérrez-Gil et al., 2018). Those populations were as below.

- 2,414 individuals of a Sardinian back-cross population, including Back-cross ewes and 10 F1 sires (Sardinian × Lacaune), and Back-cross daughters with their sires (Sechi et al., 2013).
- 1,598 individuals from Churra daughters family population, including ewes and 16 sires (Gutiérrez-Gil et al., 2018).
- 1,009 artificial insemination sires from a Lacaune grand-daughter families, including sons from 33 sires (Rupp et al., 2015).

QTL detection was implemented using linkage analyses (LA), linkage analyses combined with linkage disequilibrium (LDLA) or linkage disequilibrium (LD) with various software and programs. Latest results from the three studies are summarised in Table 2. Genome-wide significant QTL regions associated with somatic cell count were found in 21 regions in the Churra population (Gutiérrez-Gil, 2013) by LDLA on OAR1, OAR2, OAR3, OAR8, OAR11, OAR13, OAR14, OAR17, OAR18, OAR19, OAR20, OAR22 and OAR25, of which the QTL on OAR20 and 25 were confirmed by linkage analysis. The linkage analysis method highlighted an additional QTL region on OAR5. The OAR20 region validated earliest results with microsatellites markers (Gutiérrez-Gil et al., 2007) and was selected for further fine mapping. Sechi et al. (2013) identified seven regions associated with somatic cell counts in the Sardinian back-cross population on OAR3 (two regions), OAR4, OAR5, OAR12, OAR19, OAR20. Further fine mapping for the most significant regions on OAR4 and OAR20 has been projected. Finally, genome-wide significant QTL regions associated with somatic cell count were found in five regions in the Lacaune grand-daughter population on OAR3, OAR4, OAR11, OAR16, OAR23 (Rupp et al., 2015). The QTL located on OAR3 exhibited a highly significant threshold and a narrow confidence interval (< 0.5 Mb) and was further studied.

3.1.2. Fine mapping and identification of a causal mutation associated with mastitis

To fine map the most significant QTLs on OAR3 QTL in Lacaune

(Rupp et al., 2015) and OAR20 in Churra (Gutiérrez-Gil et al., 2018) ewes, authors subsequently used whole genome sequencing of three animals: a QTL sire and two of his sons with extreme phenotypes. Among candidate causal mutations, Gutiérrez-Gil et al. (2018) detected 13 variants distributed across seven immune-related genes predicted to cause an effect on protein function. The nominated genes were: *PKHDI*, *NOTCH4*, *AGER*, *MOG* and three genes orthologous to the human MHC: *ENSOARG0000009395* (*HLA-C*, *Homo sapiens*), *ENSOARG00000015002* (*HLA-B*, *Homo sapiens*) and *ENSOARG00000018075* (*BoLA*, *Bos taurus*, orthologous to human *HLA-A*).

For the OAR3 region, Rupp et al. (2015) identified a candidate SNP mutation that mapped to the coding sequence of the highly conserved gene *suppressor of cytokine signalling2* (*SOCS2*). An assay for evaluating the protein binding affinity suggested a functional knockout of the *SOCS2* gene in homozygous susceptible sheep. The authors proposed that this mutation altered actual function in the JAK/STAT/SOCS pathway and the control of the inflammatory response to infection. Genetic variants in genes involved in the JAK/STAT/SOCS pathway (*STAT5A* and *JAK2*) were previously associated with mastitis indicator traits (SCC) in Chinese Holstein cattle (Usman et al., 2014), which also supported the importance of this pathway in the determinism of the host's response to mastitis. It has to be noted that Rupp et al. (2015) have further reported a pleiotropic favorable effect of the *SOCS2* mutation on body growth and milk production. This has provided a molecular basis for the antagonism between mastitis resistance and production traits and has highlighted the need for better knowledge on such genetic variants with adverse effects to achieve optimal balancing selection.

3.1.3. Validation of QTLs in independent populations

Generally, QTLs had moderate effects; for the most significant, they explained 3% of the variance for the Churra OAR20 QTL and up to 12% for the Lacaune OAR3 QTL. The QTLs were mostly population specific, in agreement with the polygenic nature of the trait and the complexity of the resistance phenotype. When considering also less significant QTLs though, a few commonalities among populations have been identified, e.g., on OAR12 and OAR20 (Table 2). Accordingly, the 3SR consortium partners designed a public custom-designed ovine SNP chip (http://genoweb.toulouse.inra.fr/~tossier/3SR-WP3-960_snp_mastitis/) for validating the most significant QTL and the common QTL regions in independent populations at the time. The validation chip comprised 960 SNPs distributed over seven genomic regions. SNPs were selected from the OvineSNP50 beadchip or 800 K Illumina ovine chips (Nicolazzi et al., 2015) or were identified within the '3SR' project by genome resequencing.

Validation studies were performed using this custom chip in three independent dairy sheep populations: Chios (Banos et al., 2017), Lacaune and Manech Red Face ewes (Oget et al., 2019). Banos et al. (2017) confirmed significant QTL associated with four different mastitis related-traits on OAR2, OAR3, OAR5, OAR16 and OAR19; they also highlighted 14 relevant candidate genes implicated in innate immunity in these regions, based on several analyses such as a pathway and functional clustering analysis, a gene expression analysis and a transcription factor binding site analysis. The selected genes included *SOCS2*, *CTLA4*, *C6*, *C7*, *C9*, *PTGER4*, *DAB2*, *CARD6*, *OSMR*, *PLXNC1*, *IDH1*, *ICOS*, *FYB* and *LYFR* (Banos et al., 2017). Oget et al. (2019) validated four of the seven mastitis QTL regions: OAR2 (Lacaune and Manech Red Face), OAR3 (Lacaune), OAR16 (Manech Red Face), OAR19 (Lacaune). The point mutation in *SOCS2* (OAR3 QTL region), included in the 960 SNP chip, was the most significant SNP associated with three mastitis related-traits: increased SCS, increased incidence of bacterial infection and mammary abscess. This mutation was also significantly associated with increased body growth, confirming the pleiotropic effect of the *SOCS2* gene. This SNP was not present in the Manech Red Face available data. Two validated regions (OAR2 and 16) were also associated with milk production traits in both populations,

Table 2

Significant QTL influencing somatic cell scores identified by the Combined Linkage Disequilibrium and Linkage Analysis (LDLA), Linkage Disequilibrium (LD) or Linkage Analysis (LA) in three dairy sheep populations.

OAR	Position (Mb)	Confidence interval	Method	Sheep breed	Reference
1	259.05	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
2	83.10	83.1 - 83.2	LDLA	Churra	Gutiérrez-Gil et al., 2018
	140.36	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
3	185.00	185.0 - 185.1	LDLA	Churra	Gutiérrez-Gil et al., 2018
	4.40	4.3 - 4.4	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	22.65	22.6 - 22.8	LD	Lacaune	Rupp et al., 2015
	94.40	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	129.94	129.8 - 130.0	LD	Lacaune	Rupp et al., 2015
	130.10	129.4 - 131.4	LA	Lacaune	Rupp et al., 2015
	184.40	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	189.57	189.6 - 189.6	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	212.70	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	4	15.86	15.8 - 16.0	LD	Lacaune
46.99		42.8 - 56.7	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
5	68.69	68.6 - 68.9	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	77.99	74.9 - 81.5	LA	Churra	Gutiérrez-Gil et al., 2018
	102.69	102.6 - 102.8	LD	Lacaune	Rupp et al., 2015
6	71.31	71.1 - 71.5	LD	Lacaune	Rupp et al., 2015
	24.53	24.4 - 24.6	LD	Lacaune	Rupp et al., 2015
8	59.85	59.7 - 59.9	LD	Lacaune	Rupp et al., 2015
	80.60	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	82.60	81.4 - 83.5	LA	Lacaune	Rupp et al., 2015
9	27.49	27.4 - 27.6	LD	Lacaune	Rupp et al., 2015
	49.30	48.9 - 49.6	LA	Lacaune	Rupp et al., 2015
10	74.84	74.7 - 74.9	LD	Lacaune	Rupp et al., 2015
	36.70	35.8 - 41.1	LA	Lacaune	Rupp et al., 2015
11	41.18	41.0 - 41.3	LD	Lacaune	Rupp et al., 2015
	56.78	56.8 - 56.9	LDLA	Churra	Gutiérrez-Gil et al., 2018
	13.84	13.7 - 13.9	LD	Lacaune	Rupp et al., 2015
12	18.43	9.7 - 18.4	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	70.89	70.8 - 71.1	LD	Lacaune	Rupp et al., 2015
13	71.82	68.5 - 71.8	LDLA	Churra	Gutiérrez-Gil et al., 2018
	32.00	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	39.40	36.4 - 40.2	LA	Lacaune	Rupp et al., 2015
14	43.28	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	56.58	56.5 - 56.7	LD	Lacaune	Rupp et al., 2015
	5.89	5.8 - 6.0	LD	Lacaune	Rupp et al., 2015
16	36.10	35.2 - 37.0	LA	Lacaune	Rupp et al., 2015
	23.65	23.5 - 23.7	LD	Lacaune	Rupp et al., 2015
17	33.80	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	42.30	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	12.65	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
18	31.18	30.9 - 31.3	LD	Lacaune	Rupp et al., 2015
	26.17	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
19	28.60	28.4-28.7	LD	Lacaune	Rupp et al., 2015
	56.70	56.7-56.78	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	5.50	5.4-5.5	LDLA	Churra	Gutiérrez-Gil et al., 2018
20	21.52	20.9-23.8	LA	Churra	Gutiérrez-Gil et al., 2018
	23.52	22.0-28.0	LDLA	Churra	Gutiérrez-Gil et al., 2018
	23.75	13.8-25.4	LDLA	Sarda, Sarda × Lacaune	Sechi et al., 2013
	48.73	48.6-48.8	LD	Lacaune	Rupp et al., 2015
22	23.76	–	LDLA	Churra	Gutiérrez-Gil et al., 2018
	48.47	48.4-48.6	LD	Lacaune	Rupp et al., 2015
23	59.94	59.8-60.0	LD	Lacaune	Rupp et al., 2015
24	73.00	7.2-7.4	LD	Lacaune	Rupp et al., 2015
25	16.58	15.6-16.6	LDLA	Churra	Gutiérrez-Gil et al., 2018
	32.50	32.5-35.0	LDLA	Churra	Gutiérrez-Gil et al., 2018
	39.48	38.6-41.7	LA	Churra	Gutiérrez-Gil et al., 2018
26	19.57	19.5-19.7	LD	Lacaune	Rupp et al., 2015

indicating, at least in part, a genomic basis for the trade-off between milk production and mastitis resistance. These validation studies testify to the partial sharing of mastitis-related genetic mechanisms between different distant dairy sheep populations.

3.2. The input on mechanisms from gene expression data

A number of studies have investigated the host response to mastitis, using gene expression data. The studies were performed on various cell types: milk somatic cells (Bonnetfont et al., 2011), dendritic cells

(Genini et al., 2011; Toufeer et al., 2011), mammary epithelial cells (Bonnetfont et al., 2012) or mammary tissue samples (Chopra-Dewasthaly et al., 2017). Cells were exposed either to a microbial agent (*Staphylococcus aureus*, *S. epidermidis*, *Staphylococcus* spp., *Mycoplasma agalactiae*) in vivo after an intramammary challenge or to inactivated pathogens or bacterial soluble factors in vitro in cultured cells. Up to date, microarrays have mostly been used, either with ovine oligonucleotide (homologous) probes (Bonnetfont et al., 2011; 2012; Toufeer et al., 2011) or bovine cDNA (heterologous) probes (Genini et al., 2011). One publication thus far has used RNAseq profiling (Chopra-

Dewasthaly et al., 2017), but it can be anticipated, however, that future studies will use RNASeq analysis, since it gives a more exhaustive view of the transcriptome. In all studies, but the RNAseq profiling, Lacaune ewes from a divergent selection experiment based on somatic cell were used. The main outcomes of these studies were as below.

- Larger T-cell recruitment after *S. aureus* as compared to *S. epidermidis* challenge was evident.
- A significant difference was noted in the gene expression profile of milk somatic cells and dendritic cells in the genetic divergent lines, suggesting that they played a major role in genetic resistance/susceptibility, whereas mammary epithelial cells exhibited a more similar profile.
- Milk somatic cells gene profiling showed several differences between animals from the two divergent lines, suggesting a relationship between immune response and genetic resistance; this included cytokine and chemokine differential expression, a potentially more efficient neutrophil diapedesis, a more efficient clearance of the infection through TLR2 and MAPK signaling pathways, a limitation of cell proliferation and apoptosis in the resistant line.
- In dendritic cells, resistant animals showed upregulation of complement pathway genes and downregulation of the IL1R1 pathway.
- Transcriptome differences between resistant and susceptible lines were related to transcriptional activity within Milk Somatic Cells. RAR α , TP53, AHR were transcriptional factors of interest, both because of their differential expression and their interaction with a larger number of regulated genes between lines.

Interestingly, Banos et al. (2017) used crossed information from QTL detection and several public and in house datasets to propose 14 candidate genes for mastitis. This was based on the assumption that genes contributing to mastitis resistance are likely to be expressed both in the mammary gland and in immune tissues and differentially expressed in sheep bone marrow-derived macrophages in response to lipopolysaccharide administration. This analysis was completed with the study of 1 kb upstream regions of the Transcription Factor Binding Sites different in dairy and meat breeds. Further studies are required to confirm the functional impact of the candidate polymorphisms.

4. Breeding for improved mastitis resistance

4.1. Selection criteria and breeding objectives

Breeding programs in dairy sheep were first developed in the 1960s. Only a few dairy populations worldwide, mainly located in the para-Mediterranean region (Lacaune, Manech, Basco-béarnaise [France], Assaf, Latxa, Manchega, Churra [Spain], Sarda [Italy], Chios [Greece] populations) have the required size and organisation to allow development of large scale recording, genetic evaluation and breeding programmes (Barillet et al., 2001; Carta et al., 2009; Gootwine, 2011; Theodoridis et al., 2018). Milk yield has been the main selection criterion in most of those breeds during the past decades (Carta et al., 2009). Thereafter, in some breeds, a component for milk composition (Barillet et al., 2001; Gutiérrez-Gil et al., 2009) and udder morphology (Casu et al., 2006; Marie-Etancelin et al., 2006) has been added.

Following the cattle initiative in the 1980s (Heringstad et al., 2000), in several dairy sheep populations, milk somatic cell counting has been considered as a proxy for mastitis and included that in recording schemes and breeding objectives. Thus, somatic cell counting recording has been implemented in several breeds in France (Lacaune, Manech), Italy (Sarda) and Spain (Churra, Manchega, and Laxta) (Carta et al., 2009). Repeated cell counting data can be routinely recorded as part of milk recording. However, the recording cost per animal, relative to the potential income, is prohibitive for many traits other than production. Therefore, simplified schemes have been developed, with only 2–4 test days per lactation period or with sampling only in ewes in their first or/

and second parity (Rupp et al., 2002; Carta et al., 2009).

Up to date, systematic selection for somatic cell counting in addition to milk production and udder morphology is being performed only in Lacaune sheep (Rupp et al., 2002; Barillet et al., 2009). In this breed, the current relative weight for somatic cell count is 25% in the total merit index, with relative weights of somatic cell scores compared to the production of 1:2. At current selection intensities, such a combination is expected to reduce somatic cell scores by one genetic standard deviation in 10 years. Recently, a genetic evaluation for somatic cell count has been initiated in Manech Red Face sheep, updating accordingly the breeding objectives. Indirect selection for improved mastitis resistance can be also expected if udder morphology is used for selection, as this gives favorable genetic correlations between traits, as performed in Sarda sheep (Casu et al., 2006).

A divergent selection experiment based on somatic cell scores in Lacaune dairy sheep has provided evidence that selection for decreased milk somatic cell count may help to improve host's resistance to mastitis and decrease the frequency for clinical and subclinical intramammary infections (Rupp et al., 2009; Allain et al., 2010). Based on the results of approx. 200 ewes in each line selected for either high or low somatic cell scores, authors have shown a large difference between the genetic lines for somatic cell count: mean cell counts were 1,200,000 versus 280,000 cells mL⁻¹, respectively. A significant decrease of clinical mastitis, chronic clinical mastitis (as detected by the presence of parenchymal abscesses), and intramammary infections caused by various pathogens (measured by repeated milk bacteriological tests) was also observed in the low somatic cell score line when compared to the high line (Rupp et al., 2009; Allain et al., 2010). The reduced risk of mastitis when selecting for decreased somatic cell scores was further confirmed in two experiments, in which ewes from the divergent lines were challenged with *S. epidermidis* or *S. aureus* (Bonnetfont et al., 2011). The results indicated that selection for decreased somatic cell scores correlated with a better ability to control intramammary multiplication of bacteria and to limit consequences of infection and inflammation.

Minor consideration has been given to mastitis in meat sheep industry up to date. However, in a meat breed (Texel), Conington et al. (2008) have highlighted economic and welfare benefits through the reduction of antibiotic use and extra labor involved. These authors estimated that a 10% reduction in risk of developing mastitis would be leading to GBP 8.40 per ewe, or GBP 2.7 million a year to the purebred Texel population in the United Kingdom (Conington et al., 2008; McLaren et al., 2018). Additionally, the authors underlined that increased incidence of subclinical mastitis had an adverse in the body weight of lambs of affected ewes and thus in overall production output of the flock. However, given that in meat-production flocks ewes are not handled on a daily basis, as in dairy flocks, use of somatic cell count poses challenges. In view of that, McLaren et al. (2018) have suggested results of California Mastitis Test and udder conformation scoring could be useful alternative selection criteria for future genetic selection.

4.2. The contribution of genomics to breeding for mastitis resistance

4.2.1. Direct use of information from major genes

Diagnostic tools can be developed for genotyping major genes, when causal mutation or closely linked markers, are identified. As an example of a disease-related gene, testing for PrP gene associated to resistance and susceptibility to scrapie (Elsen et al., 1999), is used in breeding programs worldwide, mainly to avoid entering susceptible animals in the reproduction process. Other significant genes in sheep are *Tmem154*, associated with resistance to *Small Ruminant Lentivirus* infections (Heaton et al., 2012), and *FecL* for hyperprolificacy of ewes (Martin et al., 2014). To date, the large-scale development of such marker-assisted selection for improving mastitis resistance in various sheep populations is unlikely. Indeed, the *SOC2* gene is the only published causal mutation for mastitis resistance and the susceptible

allele was found only in the Lacaune breed (Rupp et al., 2015).

Mastitis resistance is highly polygenic with a large number of genes with small effects, which is not well suited for single gene-based selection. With the decrease of sequencing and genotyping costs and the addition of genomic studies in sheep, it is expected, though, that additional causal mutations will be available in the near future. One interesting opportunity is the development of small sets of parentage SNP or low-density chips including also major gene information, which could allow extending such diagnosis approach for various traits and population at reasonable costs.

4.2.2. Genomic selection

The objective of genomic evaluations, called by extension 'genomic selection', is to estimate the genetic value (called Estimated Breeding Value, i.e., EBV) of an individual for a given trait (e.g., mastitis resistance, i.e. somatic cell counts) based on genomic information, such as SNP markers covering the whole genome (Meuwissen et al., 2001). No information on the genes or genomic regions associated with the trait is necessary. Genomic evaluation requires a significant number of genotyped and phenotyped individuals, forming the 'reference population', in which associations between SNP and phenotypes are established. Then, using the established predictions, genomic EBVs are estimated in a population of related candidates, for which no phenotype is (yet) available. The benefit from the genomic selection is derived from the fact that genomic EBVs have higher accuracies, i.e. better predict future achieved performance, than pedigree-based EBVs, so that early selection can be efficient. Thus, genomic selection can be useful in sheep breeding to increase the genetic gain by decreasing the generation interval. This is especially true in dairy sheep because there is a long period between the dissemination of male's semen across the population and collection of phenotypes recorded in female progeny. Genomic selection in ruminants can also be useful for traits that are not measured in large-scale population as a routine, e.g., mastitis in meat sheep breeds.

A few studies have compared pedigree-based selection and selection based on genomic information for mastitis resistance (SCS trait) in the Lacaune dairy breed and have reported a gain of prediction accuracy ranging from +0.04 to +0.10 (Duchemin et al., 2012; Baloche et al., 2014). These results are lower than in cattle (+0.15) (VanRaden et al., 2009) probably mainly because reference population is smaller and the genetic diversity is higher, as proven by lower linkage disequilibrium (Baloche et al., 2014). Currently, the so-called Single Step Genomic Best Linear Unbiased Prediction (ssGBLUP) method provides the best accuracies in dairy sheep (Baloche et al., 2014); the model uses raw performances of daughters rather than ram's averages.

Future improvement of genomic selection is expected from the inclusion of major genes and QTLs in genomic evaluation models. Indeed, knowledge of genes/genomic regions with a strong effect can be added to the genome-wide SNP panel or added independently as a correlated trait, like in the Gene Content Multiple trait BLUP model (Legarra and Vitezica, 2015). Such approaches have proved to increase prediction accuracy for goat milk composition, with the inclusion of information on the α s1 casein (Carillier-Jacquin et al., 2016; Teissier et al., 2018) major genes together with the 50 K SNP chip. Application to mastitis resistance, with the inclusion of the SOCS2 gene or other QTLs, is promising.

5. Concluding remarks

In this paper, we summarized recent advances in genetic and genomic analyses for the improvement of mastitis resistance in dairy and meat-production sheep. The numerous studies on the correlations between the different traits of interest in these two sheep sectors, as well as the major new genes discovered thanks to the progress of genomic tools, pave the way for a selection of more robust individuals for sustainable breeding.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgments

Claire Oget acknowledges the support of the Agence Nationale de la Recherche (ANR) for her scholarship (Project Reidsocs, ID: ANR-16-CE20-0010). The authors are grateful to the H2020 project 'SMALL Ruminants breeding for Efficiency and Resilience' (SMARTER, https://cordis.europa.eu/project/rcn/215954_fr.html) for the support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.smallrumres.2019.05.011>.

References

- Ali, A.K.A., Shook, G.E., 1980. An optimum transformation for somatic cell concentration in milk. *J. Dairy Sci.* 63, 487–490.
- Allain, C., Aurel, M.R., Paillet, F., Portes, D., Menras, J.M., Carriere, F., Gluzel, F., Duvaillon, O., Pena-Arnaud, B., Caillat, H., 2010. La cinétique d'émission du lait et l'anatomie de la mamelle sont associées à la résistance aux mammites: résultats d'une sélection divergente de brebis sur les comptages de cellules somatiques. *Proceedings of 17èmes Rencontres autour des Recherches sur les ruminants* (Paris, France), pp. 3–4.
- Allain, C., Astruc, J.M., Portes, D., Marie-Etancelin, C., Foucras, G., Rupp, R., 2018. Genetic parameters for milk flow and relationships with bacterial infection, SCC and production traits in Lacaune dairy sheep. *Proceedings of the World Congress on Genetics Applied to Livestock Production* (Auckland, New Zealand), p. 531.
- Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.-M., Robert-Granié, C., Barillet, F., 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J. Dairy Sci.* 97, 1107–1116.
- Banos, G., Bramis, G., Bush, S.J., Clark, E.L., McCulloch, M.E.B., Smith, J., Schulze, G., Arsenos, G., Huame, D.A., Piffidi, A., 2017. The genomic architecture of mastitis resistance in dairy sheep. *BMC Genom.* 18, 624.
- Barillet, F., 2007. Genetic improvement for dairy production in sheep and goats. *Small Rumin. Res.* 70, 60–75.
- Barillet, F., Rupp, R., Mignon-Grasteau, S., Astruc, J.-M., Jacquin, M., 2001. Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genet. Sel. Evol.* 33, 397–415.
- Barillet, F., Astruc, J.M., Lagriffoul, G., Aguerre, X., Bonaiti, B., 2009. Selecting Milk Composition and Mastitis Resistance by Using a Part Lactation Sampling Design in French Manech Red Faced Dairy Sheep Breed. *ICAR Technical Series* 129–135.
- Baro, J.A., Carriedo, J.A., San Primitivo, F., 1994. Genetic parameters of test day measures for somatic cell count, milk yield, and protein percentage of milking ewes. *J. Dairy Sci.* 77, 2658–2662.
- Bergonier, D., de Crémoux, R., Rupp, R., Lagriffoul, G., Berthelot, X., 2003. Mastitis of dairy small ruminants. *Vet. Res.* 34, 689–716.
- Bonnefont, C.M., Toufeer, M., Gaubet, C., Foulon, E., Tasca, C., Aurel, M.-R., Bergonier, D., Boullier, S., Robert-Granié, C., Foucras, G., Rupp, R., 2011. Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. *BMC Genom.* 12, 208.
- Bonnefont, C.M.D., Rainard, P., Cunha, P., Gilbert, F.B., Toufeer, M., Aurel, M.-R., Rupp, R., Foucras, G., 2012. Genetic susceptibility to *S. aureus* mastitis in sheep: differential expression of mammary epithelial cells in response to live bacteria or supernatant. *Physiol. Genom.* 44, 403–416.
- Bouvier-Muller, J., Allain, C., Enjalbert, F., Tabouret, G., Portes, D., Caubet, C., Tasca, C., Foucras, G., Rupp, R., 2016. Response to dietary-induced energy restriction in dairy sheep divergently selected for resistance or susceptibility to mastitis. *J. Dairy Sci.* 99, 480–492.
- Bouvier-Muller, J., Allain, C., Enjalbert, F., Farizon, Y., Portes, D., Foucras, G., Rupp, R., 2018. Somatic cell count-based selection reduces susceptibility to energy shortage during early lactation in a sheep model. *J. Dairy Sci.* 101, 2248–2259.
- Carillier-Jacquin, C., Larroque, H., Robert-Granié, C., 2016. Including α s1 casein gene information in genomic evaluations of French dairy goats. *Genet. Sel. Evol.* 48, 54.
- Carta, A., Casu, S., Salaris, S., 2009. Invited review: current state of genetic improvement in dairy sheep. *J. Dairy Sci.* 92, 5814–5833.
- Casu, S., Pernazza, L., Carta, A., 2006. Feasibility of a linear scoring method of udder morphology for the selection scheme of Sardinian sheep. *J. Dairy Sci.* 89, 2200–2209.
- Casu, S., Sechi, S., Salaris, S.L., Carta, A., 2010. Phenotypic and genetic relationships between udder morphology and udder health in dairy ewes. *Small Rumin. Res.* 88, 77–83.
- Chopra-Dewasthaly, R., Korb, M., Brunthaler, R., Ertl, R., 2017. Comprehensive RNA-seq profiling to evaluate the sheep mammary gland transcriptome in response to experimental *Mycoplasma agalactiae* infection. *PLoS One* 12, e0170015.
- Conington, J., Cao, G., Stott, A., Bünger, L., 2008. Breeding for resistance to mastitis in United Kingdom sheep, a review and economic appraisal. *Vet. Rec.* 162, 369–376.

- Crawley, A.M., Mallard, B., Wilkie, B.N., 2005. Genetic selection for high and low immune response in pigs: effects on immunoglobulin isotype expression. *Vet. Immunol. Immunopathol.* 108, 71–76.
- Davies, G., Genini, S., Bishop, S.C., Giuffra, E., 2009. An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. *Animal* 3, 415–436.
- De La Fuente, L.F., Gonzalo, C., Sánchez, J., Rodríguez, R., Carriedo, J., Primitivo, F., 2011. Genetic parameters of the linear body conformation traits and genetic correlations with udder traits, milk yield and composition, and somatic cell count in dairy ewes. *Can. J. Anim. Sci.* 91, 585–591.
- Detilleux, J.C., 2002. Genetic factors affecting susceptibility of dairy cows to udder pathogens. *Vet. Immunol. Immunopathol.* 88, 103–110.
- Duchemin, S.I., Colombani, C., Legarra, A., Baloché, G., Larroque, H., Astruc, J.-M., Barillet, F., Robert-Granié, C., Manfredi, E., 2012. Genomic selection in the French Lacaune dairy sheep breed. *J. Dairy Sci.* 95, 2723–2733.
- El-Saied, U.M., Carriedo, J.A., De La Fuente, L.F., San Primitivo, F., 1999. Genetic parameters of lactation cell counts and milk and protein yields in dairy ewes. *J. Dairy Sci.* 82, 639–644.
- Elsen, J.M., Amigues, Y., Schelcher, F., Ducrocq, V., Andreoletti, O., Eycheur, F., Khang, J.V., Poivey, C., J.P., Lantier, F., Laplanche, J.L., 1999. Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. *Arch. Virol.* 144, 431–445.
- Emanuelson, U., Philipsson, J., 1984. Studies on somatic cell counts in milk from Swedish dairy cows. *Acta Agric. Scand.* 34, 45–53.
- Emanuelson, U., Danell, B., Philipsson, J., 1988. Genetic parameters for clinical mastitis, somatic cell counts, and milk production estimated by multiple-trait restricted maximum likelihood. *J. Dairy Sci.* 71, 467–476.
- Genini, S., Badaoui, B., Sclep, G., Bishop, S.C., Waddington, D., Pinard van der Laan, M.-H., Klopp, C., Cabau, C., Seyfert, H.-M., Petzl, W., Jensen, K., Glass, E.J., de Greeff, A., Smith, H.E., Smits, M.A., Olsaker, I., Buman, G.M., Pisoni, G., Moroni, P., Castiglioni, B., Cremonesi, P., Del Corvo, M., Foulon, E., Fourcra, G., Rupp, R., Giuffra, E., 2011. Strengthening insights into host responses to mastitis infection in ruminants by combining heterogeneous microarray data sources. *BMC Genom.* 12, 225.
- Gonzalo, C., Ariznabarreta, A., Othmane, M.H., Carriedo, J.A., Fuente, L.F.D.L., Primitivo, F.S., 2003. Genetic parameters of somatic cell count in dairy sheep considering the type of mammary pathogen effect. *J. Anim. Breed. Genet.* 120, 282–287.
- Gootwine, E., 2011. Mini review: breeding Awassi and Assaf sheep for diverse management conditions. *Trop. Anim. Health Prod.* 43, 1289–1296.
- Gutiérrez-Gil, B., 2013. Detection of QTL influencing somatic cell score in churra sheep employing the OvineSNP50 beadchip. *Book of Abstracts of the 64th Annual Meeting of the European Federation of Animal Science (Nantes, France)*. p. 601.
- Gutiérrez-Gil, B., El-Zarei, M.F., Bayón, Y., Álvarez, L., de la Fuente, L.F., Primitivo, F.S., Arranz, J.J., 2007. Detection of quantitative trait loci influencing somatic cell score in Spanish Churra sheep. *J. Dairy Sci.* 90, 422–426.
- Gutiérrez-Gil, B., Esteban-Blanco, C., Suárez-Vega, A., Arranz, J.J., 2018. Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50k SNP-Chip and whole genome sequencing. *J. Dairy Sci.* 101, 9072–9088.
- Gutiérrez-Gil, B., El-Zarei, M.F., Álvarez, L., Bayón, Y., De La Fuente, L.F., San Primitivo, F., Arranz, J.-J., 2009. Quantitative trait loci underlying milk production traits in sheep. *Anim. Genet.* 40, 423–434.
- Hamann, H., Horstlick, A., Wessels, A., Distl, O., 2004. Estimation of genetic parameters for test day milk production, somatic cell score and litter size at birth in East Friesian ewes. *Liv. Prod. Sci.* 87, 153–160.
- Heaton, M.P., Clawson, M.L., Chitko-Mckown, C.G., Leymaster, K.A., Smith, T.P.L., Harhay, G.P., White, S.N., Herrmann-Hoesing, L.M., Mousel, M.R., Lewis, G.S., Kalbfleisch, T.S., Keen, J.E., Laegreid, W.W., 2012. Reduced lentivirus susceptibility in sheep with TMEM154 mutations. *PLoS Genet.* 8, e1002467.
- Heringstad, B., Klemetsdal, G., Ruane, J., 2000. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. *Liv. Prod. Sci.* 64, 95–106.
- Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., Muzny, D.M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W.C., Hourlier, T., Aken, B.L., Searle, S.M.J., Adelson, D.L., Bian, C., Cam, G.R., Chen, Y., Cheng, S., DeSilva, U., Dixon, K., Dong, Y., Fan, G., Franklin, L.R., Fu, S., Guan, R., Highland, M.A., Holder, M.E., Huang, G., Ingham, A.B., Jhangiani, S.N., Kalra, D., Kovar, C.L., Lee, S.L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, Wenliang, Wei, B., White, S.N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J.B., Kristensen, K., Gibbs, R.A., Plicek, P., Warkup, C.C., Jones, H.E., Oddy, V.H., Nicholas, F.W., McEwan, J.C., Kijas, J., Wang, J., Worley, K.C., Archibald, A.L., Cockett, N., Xu, X., Wang, Wen, Dalrymple, B.P., 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168–1173.
- Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V.H., Nicholas, F.W., Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4, e4668.
- Legarra, A., Ugarte, E., 2005. Genetic parameters of udder traits, somatic cell score, and milk yield in Latsxa sheep. *J. Dairy Sci.* 88, 2238–2245.
- Legarra, A., Vitezica, Z.G., 2015. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. *Genet. Sel. Evol.* 47, 89.
- Marie-Etancelin, C., Manfredi, E., Aurel, M.-R., Pailler, E., Arhainx, J., Ricard, E., Lagriffoul, G., Guillouet, P., Bibé, B., Barillet, F., 2006. Genetic analysis of milking ability in Lacaune dairy ewes. *Genet. Sel. Evol.* 38, 183–200.
- Martin, P., Raoul, J., Bodin, L., 2014. Effects of the FecL major gene in the Lacaune meat sheep population. *Genet. Sel. Evol.* 46, 48.
- McLaren, A., Kaseja, K., Yates, J., Mucha, S., Lambe, N.R., Conington, J., 2018. New mastitis phenotypes suitable for genomic selection in meat sheep and their genetic relationships with udder conformation and lamb live weights. *Animal* 12, 2470–2479.
- Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Mouton, D., Bouthillier, Y., Mevel, J.C., Bizzio, G., 1984. Genetic selection for antibody responsiveness in mice: further evidence for inverse modification of macrophage catabolic activity without alteration of the expression of T-cell-mediated immunity. *Ann. Immunol. (Paris)* 135D, 173–186.
- Mrode, R., Swanson, G., 1996. Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle. *Anim. Breed. Abstracts* 64, 847–857.
- Nicolazzi, E.L., Caprera, A., Nazzicari, N., Cozzi, P., Strozzi, F., Lawley, C., Pirani, A., Soans, C., Brew, F., Jorjani, H., Evans, G., Simpson, B., Tosser-Klopp, G., Brauning, R., Williams, J.L., Stella, A., 2015. SNPchip v3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genom.* 16, 283.
- O'Brien, A.C., McHugh, N., Wall, E., Pabion, T., McDermott, R., Randles, S., Fair, S., Berry, D.P., 2017. Genetic parameters for lameness, mastitis and dagginess in a multi-breed sheep population. *Animal* 11, 911–919.
- Oget, C., Allain, C., Portes, D., Fourcra, G., Stella, A., Astruc, J.-M., Sarry, J., Tosser-Klopp, G., Rupp, R., 2019. A validation study of loci associated with mastitis resistance in two French dairy sheep breeds. *Genet. Sel. Evol.* 51, 5.
- Olsen, H.G., Knutsen, T.M., Lewandowska-Sabat, A.M., Grove, H., Nome, T., Svendsen, M., Arnyasi, M., Sodeland, M., Sundsaasen, K.K., Dahl, S.R., Heringstad, B., Hansen, H.H., Olsaker, I., Kent, M.P., Lien, S., 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. *Genet. Sel. Evol.* 48, 79.
- Othmane, M.H., Carriedo, J.A., San Primitivo, F., De la Fuente, L.F., 2002a. Genetic parameters for lactation traits of milking ewes: protein content and composition, fat, somatic cells and individual laboratory cheese yield. *Genet. Sel. Evol.* 34, 581–596.
- Othmane, M.H., De La Fuente, L.F., Carriedo, J.A., San Primitivo, F., 2002b. Heritability and genetic correlations of test day milk yield and composition, individual laboratory cheese yield, and somatic cell count for dairy ewes. *J. Dairy Sci.* 85, 2692–2698.
- Pinard van der Laan, M.-H., 2002. Immune modulation: the genetic approach. *Vet. Immunol. Immunopathol.* 87, 199–205.
- Riggio, V., Portolano, B., 2015. Genetic selection for reduced somatic cell counts in sheep milk: a review. *Small Rumin. Res.* 126, 33–42.
- Riggio, V., Finocchiaro, R., van Kaam, J.B.C.H.M., Portolano, B., Bovenhuis, H., 2007. Genetic parameters for milk somatic cell score and relationships with production traits in primiparous dairy sheep. *J. Dairy Sci.* 90, 1998–2003.
- Riggio, V., Portolano, B., Bovenhuis, H., Bishop, S.C., 2010. Genetic parameters for somatic cell score according to udder infection status in Valle del Belice dairy sheep and impact of imperfect diagnosis of infection. *Genet. Sel. Evol.* 42, 30.
- Rupp, R., Boichard, D., 2003. Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34, 671–688.
- Rupp, R., Fourcra, G., 2010. Genetics of mastitis in dairy ruminants. In: Bishop, S.C., Axford, R.F.E., Nicholas, F.W., Owen, J.B. (Eds.), *Breeding for Disease Resistance in Farm Animals*. CAB, Wallingford, pp. 183–212.
- Rupp, R., Boichard, D., Barhat, A., Astruc, J.M., Lagriffoul, G., Barillet, F., 2002. Selection for mastitis resistance in French dairy sheep. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production (Montpellier, France)*. pp. 119–122.
- Rupp, R., Lagriffoul, G., Astruc, J.M., Barillet, F., 2003. Genetic parameters for milk somatic cell scores and relationships with production traits in French Lacaune dairy sheep. *J. Dairy Sci.* 86, 1476–1481.
- Rupp, R., Bergeron, D., Dion, S., Hygonenq, M.C., Aurel, M.R., Robert-Granié, C., Fourcra, G., 2009. Response to somatic cell count-based selection for mastitis resistance in a divergent selection experiment in sheep. *J. Dairy Sci.* 92, 1203–1219.
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tascia, C., Ligat, L., Portes, D., Woloszyn, F., Bouchez, O., Tabouret, G., Lebastard, M., Caubet, C., Fourcra, G., Tosser-Klopp, G., 2015. A point mutation in suppressor of cytokine signalling 2 (Soes2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet.* 11, e1005629.
- Sechi, S., Casu, S., Castala, M., Congiu, G.B., Miari, S., Mulas, G., Salaris, S., Sechi, T., Usai, M.G., Ligios, C., Fourcra, G., Carta, A., 2013. Genome-wide association analysis of resistance to paratuberculosis and mastitis in dairy sheep. *Book of Abstracts of the 64th Annual Meeting of the European Federation of Animal Science (Nantes, France)*. p. 597.
- Serrano, M., Pérez-Guzmán, M.D., Montoro, V., Jurado, J.J., 2003. Genetic analysis of somatic cell count and milk traits in Manchega ewes: mean lactation and test-day approaches. *Liv. Prod. Sci.* 84, 1–10.
- Teissier, M., Larroque, H., Robert-Granié, C., 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genet. Sel. Evol.* 50, 31.
- Theodoridis, A., Ragkos, A., Rose, G., Roustemis, D., Arsenos, G., 2018. Defining the breeding goal for a sheep breed including production and functional traits using market data. *Animal* 12, 1508–1515.
- Tolone, M., Riggio, V., Portolano, B., 2013. Estimation of genetic and phenotypic parameters for bacteriological status of the udder, somatic cell score, and milk yield in dairy sheep using a threshold animal model. *Liv. Sci.* 151, 134–139.
- Tolone, M., Larroque, C., Yáñez, J.M., Newman, S., Sardina, M.T., Portolano, B., 2016. Assessment of genetic variation for pathogen-specific mastitis resistance in Valle del Belice dairy sheep. *BMC Vet. Res.* 12, 158.

- Toufeer, M., Bonnefont, C.M.D., Foulon, E., Caubet, C., Tasca, C., Aurel, M.-R., Robert-Granié, C., Rupp, R., Foucras, G., 2011. Gene expression profiling of dendritic cells reveals important mechanisms associated with predisposition to staphylococcus infections. *PLoS One* 6, e22147.
- Traoré, I., Péiffer, H., Prévot, F., Grisez, C., Bergeaud, J.P., Rupp, R., Aurel, M.R., Foucras, G., Jacquet, P., 2008. Effect of genetic selection for mastitis resistance in the Lacune breed of sheep on the response to *Haemonchus contortus* infection. Proceedings of the Xth European Multicolloquium of Parasitology (Paris, France), p. 88.
- Usman, T., Yu, Y., Liu, C., Wang, X., Zhang, Q., Wang, Y., 2014. Genetic effects of single nucleotide polymorphisms in JAK2 and STAT5A genes on susceptibility of Chinese Holsteins to mastitis. *Molec. Biol. Rep.* 41, 8293–8301.
- VanRaden, P.M., Van Tassel, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., Schenkel, F.S., 2009. Invited review: reliability of genomic predictions for north american holstein bulls. *J. Dairy Sci.* 92, 16–24.

IV. Conclusions du Chapitre I. et questions scientifiques soulevées

Dans ce chapitre introductif, nous avons vu les principes des schémas de sélection en élevage rendus possibles par l'existence d'un déterminisme génétique sur les caractères d'intérêt de l'objectif de sélection. En effet, les variabilités génétiques au niveau des *loci* du génome qui déterminent ces caractères d'intérêt, telles que les mutations, sont à l'origine de la variabilité phénotypique observable et sélectionnable dans la population. En ovin, et comme dans toutes les espèces d'élevage, les caractères sélectionnés sont majoritairement déterminés par de multiples *loci* du génome à effets faibles qui peuvent être couplés à des gènes majeurs à effets forts. La mutation ponctuelle à l'étude dans cette thèse (R96C) est située dans un gène majeur pléiotrope appelé *SOCS2*, qui semble être capable d'influencer plusieurs caractères tels que le processus d'inflammation mammaire, la production laitière ou encore la croissance. La prise en compte de l'information de cette mutation en sélection paraît alors être un point important pour l'amélioration efficace de l'espèce. Une première question scientifique soulevée par ce constat est la suivante : quels sont les moyens disponibles pour la prise en compte de cette mutation pléiotrope dans le contexte de la sélection génomique, mise en place aujourd'hui en brebis Lacaune sur plusieurs caractères (production, santé, et morphologie mammaire) ? Pour y répondre, différents modèles d'évaluation génétique et génomique sur les neuf caractères en sélection vont être testés ([Chapitre II](#)).

Dans un second temps, nous avons pu constater que ce gène pléiotrope codait pour la protéine *SOCS-2*, une protéine intervenant dans la voie de transmission JAK/STAT qui est la principale voie de transmission du signal cellulaire pour une vaste palette de cytokines et de facteurs de croissance chez les mammifères. Cette voie de transmission complexe et actrice dans de nombreuses fonctions biologiques est en effet régulée par un grand nombre de molécules dont notamment les protéines inhibitrices de la famille *SOCS*. La régulation négative de la voie JAK/STAT par les *SOCS* est réalisée principalement *via* leur capacité de liaison avec les autres molécules de la voie tels que le récepteur lui-même, les protéines JAK, ou encore les autres molécules régulatrices. Cependant, les mécanismes sous-jacents restent encore incertains, et encore plus particulièrement pour *SOCS-2*, une des protéines peu étudiées au sein de cette famille. Une mutation telle que celle étudiée dans cette thèse, localisée dans le domaine de liaison SH2, devrait pouvoir permettre d'éclairer cet aspect. Une deuxième question scientifique soulevée par la multitude de fonctions biologiques pouvant potentiellement être régulées par *SOCS-2* est celle de la confirmation, la validation et la découverte de l'association de cette mutation aux différents caractères d'intérêt pour la filière ovine, au-delà des seuls caractères pris en compte en sélection. Pour ce faire, différentes approches statistiques vont être

utilisées afin de quantifier l'association de la mutation à des caractères en lien avec la santé, la production, la croissance et la reproduction ([Chapitre III](#)).

Enfin, nous avons pu replacer cette mutation, découverte en ovin en raison de son effet très fort et négatif sur l'inflammation mammaire, dans le contexte des études génétiques et génomiques sur la résistance aux mammites en ovin, un caractère complexe. En effet, ce caractère de faible héritabilité semble avoir un déterminisme partagé au sein des différentes races et dispersé sur plusieurs QTL dans le tout le génome. Nous avons également mis en lumière les relations génétiques complexes entre les différents caractères d'intérêt en ovin, à prendre en compte en sélection. Une troisième question scientifique a été soulevée par l'effet fort de la mutation sur ce caractère d'intérêt majeur de santé mammaire : quels sont les mécanismes moléculaires biologiques sous-jacents de la réponse à l'infection et différentiels entre les individus porteurs de la mutation et ceux non porteurs ? Cette question sera traitée à l'aide d'un protocole expérimental où des brebis portant les génotypes *SOCS2* extrêmes seront infectées et suivies finement pour un grand nombre de phénotypes dont fait partie l'expression transcriptomique sanguine.

Chapitre II. Étude prospective de la prise en compte de la mutation *SOCS2* dans le contexte de la sélection génomique

I. Objectifs et démarches

La mutation *SOCS2* influençant plusieurs caractères d'intérêt en ovin, nous nous sommes intéressés à son effet sur l'ensemble des neuf caractères en sélection en brebis Lacaune en France : cinq caractères de production laitière, un caractère de santé (SCS) et trois caractères de morphologie mammaire. Dans cette race, la sélection génomique est déjà implémentée (Astruc *et al.*, 2012) et utilise des données de génotypages de moyenne densité (puce 54K ovine) des mâles améliorateurs des stations de testage. Cependant, la méthode utilisée ne permet pas d'apporter, *a priori*, l'information de cette mutation majeure qui n'est pas présente dans le panel de SNP de la puce 54K. Le génotypage de la mutation *SOCS2* était disponible pour un certain nombre d'individus et nous avons donc pu imputer les génotypes *SOCS2* manquants à l'ensemble de la population génotypée avec la puce 54K.

L'objectif principal de ce chapitre est donc d'évaluer si la prise en compte de la mutation causale du gène *SOCS2*, dans la mesure où elle est connue, permet d'améliorer la sélection. Il s'agit donc de tester différentes méthodes permettant d'inclure la mutation dans l'évaluation génomique actuelle. Cette étude a reposé sur la comparaison des précisions des valeurs génétiques prédites par les différents modèles, et est inspiré du travail réalisé par Marc Teissier (Teissier *et al.*, 2018 ; Teissier *et al.*, 2019) en caprin. Un autre objectif de ce travail était de quantifier les effets génétiques dus à cette mutation sur les différents caractères en sélection.

Ce chapitre sera présenté sous forme d'un article scientifique accepté pour publication le 28 août 2019 dans la revue *BMC Genomics*. Les figures additionnelles (Figures S1 à S5) de cet article sont présentes en annexe (respectivement Annexes 2 à 6).

II. Résumé de l'article scientifique

II.1. Introduction

Dans cette étude, la précision de diverses méthodes d'évaluation génomique, permettant l'introduction de l'effet d'une mutation causale, a été évaluée. Nous avons pris comme exemple d'étude une mutation ponctuelle dans le gène *SOCS2* chez la brebis laitière qui est associée à la croissance, à la production laitière et au SCS, un caractère lié aux mammites.

II.2. Méthodes

Le jeu de données étudié comprenait 1 503 148 femelles phénotypées et 9 844 individus génotypés avec la puce ovine Illumina de 54K SNP. Le SNP *SOCS2* était génotypé pour 4 297 animaux et a été imputé pour les 9 844 animaux. Nous avons réalisé des évaluations génétiques pour les caractères suivants : production laitière, quantités de matières et taux protéique et butyreux, SCS, ainsi que trois caractères de morphologie mammaire. Pour chaque caractère, les valeurs génétiques (EBVs, *Estimated Breeding Values*) et leur précision ont été estimées par des approches *single-step* (« en une seule étape ») utilisant des phénotypes femelles. Dans un premier temps, six méthodes ont été testées et comparées : i) BLUP pedigree, ii) BLUP génomique (ssGBLUP), qui utilise les 54K SNP de la puce, et iii) quatre méthodes ssGBLUP pondérées (WssGBLUP). Les méthodes WssGBLUP attribuent des poids aux SNP en fonction de leur effet sur le caractère. Ensuite, afin d'inclure la mutation causale, les cinq méthodes génomiques ont de nouveau été testées après l'ajout de *SOCS2* comme SNP parmi les marqueurs de la puce. Enfin, nous avons également testé la méthode *Gene Content* qui considère le génotype *SOCS2* comme un caractère dans un modèle multi-caractères.

II.3. Résultats

Nous avons obtenu un gain de précision des EBVs de 14,03% entre les méthodes BLUP pedigree et ssGBLUP, et de 3,99% entre les méthodes ssGBLUP et WssGBLUP. De plus, l'ajout du SNP *SOCS2* dans les méthodes ssGBLUP a conduit à un gain supplémentaire de 0,26%. Le fait de mettre l'accent sur les SNP dans les régions ayant un fort effet sur les caractères, sans connaissance *a priori* de ces régions QTL, pour construire la matrice de parenté a donc permis une meilleure précision des EBVs. Lorsqu'aucune donnée de génotypage n'a été utilisée dans les modèles, la méthode du *Gene Content* a permis de prendre en compte efficacement l'information de génotypage partiel qu'est le génotype *SOCS2*, car la précision des EBVs a été augmentée de 6,25 % par rapport au BLUP pedigree. Cette méthode a également permis de dissocier la composante génétique due au gène majeur (*SOCS2*) de la composante polygénique restante.

II.4. Conclusion

L'information de mutations causales ayant un effet modéré à fort peut être capturée avec des puces à SNP de moyenne densité en appliquant des méthodes d'évaluation génomique appropriées. La méthode du *Gene Content* est efficace pour tenir compte des mutations causales dans des populations pour lesquelles des données de génotypage à l'échelle du génome ne sont pas disponibles.

II.5. Correctif sans conséquence pour l'article (Oget *et al.*, 2019b)

Une faute de frappe s'est glissée dans le premier paragraphe de la partie « *SOCS2 genotypes and imputation of missing data* » de la section « *Methods* » de l'article scientifique (Oget *et al.*, 2019b). Les individus génotypés pour la mutation *SOCS2* (tests KASPar™) sont de millésimes de naissance entre 2002 et 2013 et non 2002 et 2003 comme indiqué dans l'article. Cette erreur est sans conséquence sur les résultats de l'étude ni sur son interprétation car il s'agit d'une faute de frappe et nous avons l'information véridique pour le design de l'analyse d'imputation.

III. Article scientifique (Oget *et al.*, 2019b)

RESEARCH ARTICLE

Open Access



Alternative methods improve the accuracy of genomic prediction using information from a causal point mutation in a dairy sheep model

Claire Oget^{1*}, Marc Teissier¹, Jean-Michel Astruc², Gwenola Tosser-Klopp^{1†} and Rachel Rupp^{1†}

Abstract

Background: Genomic evaluation is usually based on a set of markers assumed to be linked with causal mutations. Selection and precise management of major genes and the remaining polygenic component might be improved by including causal polymorphisms in the evaluation models. In this study, various methods involving a known mutation were used to estimate prediction accuracy. The *SOCS2* gene, which influences body growth, milk production and somatic cell scores, a proxy for mastitis, was studied as an example in dairy sheep.

Methods: The data comprised 1,503,148 phenotypes and 9844 54K SNPs genotypes. The *SOCS2* SNP was genotyped for 4297 animals and imputed in the above 9844 animals. Breeding values and their accuracies were estimated for each of nine traits by using single-step approaches. Pedigree-based BLUP, single-step genomic BLUP (ssGBLUP) involving the 54K ovine SNPs chip, and four weighted ssGBLUP (WssGBLUP) methods were compared. In WssGBLUP methods, weights are assigned to SNPs depending on their effect on the trait. The ssGBLUP and WssGBLUP methods were again tested after including the *SOCS2* causal mutation as a SNP. Finally, the Gene Content approach was tested, which uses a multiple-trait model that considers the *SOCS2* genotype as a trait.

Results: EBV accuracies were increased by 14.03% between the pedigree-based BLUP and ssGBLUP methods and by 3.99% between ssGBLUP and WssGBLUP. Adding the *SOCS2* SNP to ssGBLUP methods led to an average gain of 0.26%. Construction of the kinship matrix and estimation of breeding values was generally improved by placing emphasis on SNPs in regions with a strong effect on traits. In the absence of chip data, the Gene Content method, compared to pedigree-based BLUP, efficiently accounted for partial genotyping information on *SOCS2* as accuracy was increased by 6.25%. This method also allowed dissociation of the genetic component due to the major gene from the remaining polygenic component.

Conclusions: Causal mutations with a moderate to strong effect can be captured with conventional SNP chips by applying appropriate genomic evaluation methods. The Gene Content method provides an efficient way to account for causal mutations in populations lacking genome-wide genotyping.

Keywords: Genomics, Genomic evaluation, Genome-wide association study, Dairy sheep, Causal mutation

* Correspondence: claire.oget@inra.fr

[†]Gwenola Tosser-Klopp and Rachel Rupp contributed equally to this work.

¹GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

By estimating genetic parameters, such as the heritability of a given trait, individuals could be selected according to their genetic value, based on the Estimated Breeding Values (EBVs) for that trait, and the whole species might be genetically improved for traits such as production, health, morphology, etc. Schemes were therefore set up to select improved males and use their semen on livestock breeding farms. The genetic architecture of traits of interest in livestock species has been widely studied since the 1920s [1]. Such traits can be governed by genes with small effects but also by large Quantitative Trait *Loci* (QTLs) or major genes. Studies of this complex architecture have been facilitated by new technologies and molecular markers such as microsatellites or single nucleotide polymorphisms (SNPs) which make it possible to detect the regions of the genome responsible for genetic variation and measure their respective effects [1].

Genetic selection methods were initially based on pedigree approaches and the method first employed to estimate breeding values was the Best Linear Unbiased Prediction method (BLUP) [2]. Approaches based on SNP chips were then quickly developed from the 1990s onwards [3–6]. These approaches allowed EBVs to be estimated from pedigree information, from genotyping data about a proportion of the population (males in testing stations for example), and from performance data. Performance data could be based on means of progeny performance, e.g. Daughter Yield Deviations (DYD), as in the two-step pedigree-based BLUP [2] or Genomic BLUP (GBLUP) [3, 4] approaches. More recently, methods to directly use raw phenotypes of non-genotyped individuals in the so-called single-step GBLUP approach (ssGBLUP) were developed [5, 6]. A few studies showed that the prediction accuracy of evaluations could be increased by using ssGBLUP, rather than two-step pedigree-based BLUP or GBLUP approaches [7–9].

Two studies, in the same dairy Lacaune breed sheep population investigated here, resulted in the development of genetic evaluation models based on molecular markers [10, 11]. Duchemin et al. (2012) [10], after comparing the BLUP, Bayes C π , Partial Least Squares (PLS), and sparse PLS methods, reported that depending on the trait and compared to the BLUP method, EBV accuracies could be increased by 18 to 25% by including markers in the models, with minor differences between the genomic approaches. Baloché et al. (2014) [11] then adopted BLUP-like methods to implement a single-step model in the evaluation and compared pseudo-BLUP and pseudo-ssGBLUP (using all rams and their DYDs in both methods), and regular ssGBLUP (using individual phenotypes and pedigree in an animal model), and obtained the best results with regular ssGBLUP. In 2015, the ssGBLUP approach was therefore implemented in the French official genetic evaluations of Lacaune sheep

[12] and is used as a reference method in this study. However, the previously tested methods, and the one currently used in official evaluations, do not allow a higher weight to be assigned to markers in QTL regions or to a major gene such as the *SOCS2* gene, which influences many traits due to the mutation present in this population.

In the ssGBLUP approach, all SNPs are given the same weight during construction of the relationship matrix. Methods have since been developed to assign more weight to markers that are more strongly associated with the trait under study [13, 14] or to a major gene influencing the trait, in a multi-trait approach (called Gene Content) [15]. These methods have been tested in goats and have been shown to improve evaluation accuracy [16–18].

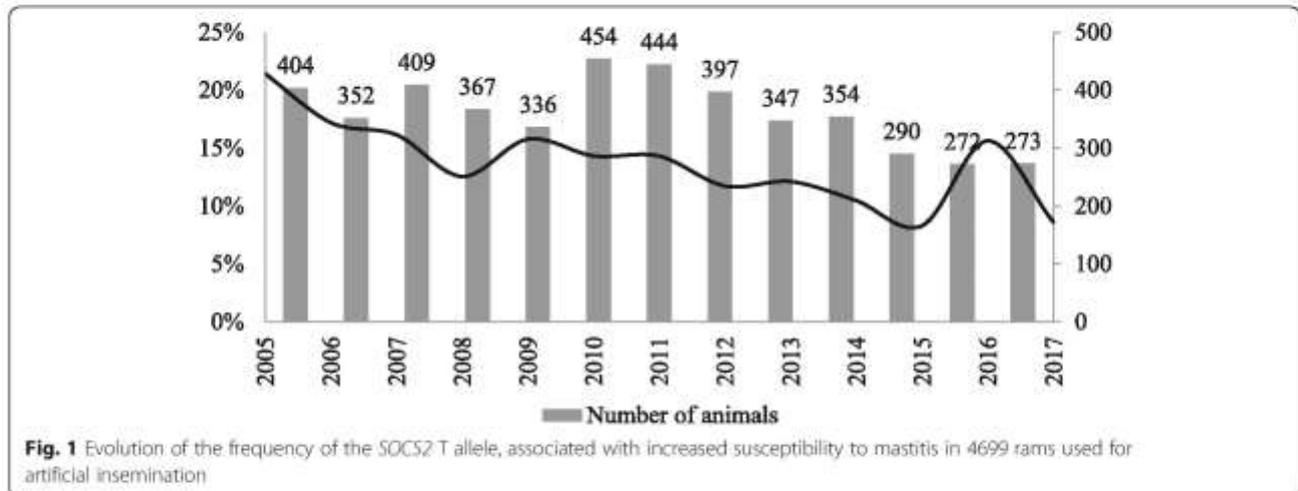
In this study, we chose the causal mutation for mastitis resistance characterized in a dairy sheep association study: namely the R96C point mutation in the *SOCS2* gene (suppressor of cytokine signaling 2) [19]. This mutation, which consists of a modification in a single base pair (substitution of an allelic base C into an allelic base T), introduces a SNP at this *locus* and modifies the affinity of the protein for its ligand. Rupp et al. (2015) [19] showed that this mutation, with a Minor Allele Frequency (MAF) of 21.7% in the population (468 rams from testing stations), was strongly associated (deteriorated health) with the Somatic Cell Count (SCC) trait, considered as a proxy for mastitis (i.e. it explained 12% of the genetic variance of this trait), but was also favourably associated with size, weight and, to a lesser extent, with milk yield traits. This pleiotropic gene therefore seemed a good candidate for testing several approaches to exploit information about QTL or causal mutations in evaluation models.

Thus the objectives of this study were: (i) to test evaluation methods that allowed inclusion of information about a causal mutation, using the example of the dairy sheep *SOCS2* gene point mutation, and (ii) to analyze the effect of applying different methods to utilize this additional information on prediction accuracies and on EBV trends over time.

Results

Imputation of *SOCS2* genotypes

We obtained a Concordance Rate (CR) of 0.988 for imputation of the *SOCS2* genotypes, i.e. 33 imputation errors among the 1432 individuals in the imputation validation population. The result of this imputation provided us with the *SOCS2* genotypes for the entire genotyped population, with a MAF of 0.14 for the mutated T allele associated with higher susceptibility to mastitis. The MAF trend in the population of 4699 AI rams is shown in Fig. 1. The MAF decreased from 0.21 in 2005 to 0.09 in 2017.



Linkage Disequilibrium (LD)

The map coverage of the chip attained 2444 Mb and the mean SNP interval was 0.064 Mb. The mean r^2 were 0.25, 0.16, 0.12, 0.09, and 0.08 for a distance between pairs of SNPs of < 0.02 Mb, [0.02–0.04 Mb], [0.04–0.06 Mb], [0.06–0.08 Mb], and [0.08–0.10 Mb], respectively, and < 0.08 for all the other distance categories. A visualization of r^2 according to distance between SNP is provided in Additional file 1: Figure S1.

The r^2 measure of LD between the 40 markers closest to *SOCS2* is represented in Additional file 1: Figure S2. The LD of *SOCS2* with the other SNP markers ranged from zero to 0.47 with OAR3_138135461.1, which was 0.229 Mb pairs away from *SOCS2*. The average LD between the *SOCS2* SNP and the 10 previous SNPs on the chip was 0.17. The category containing the distance between the *SOCS2* SNP and the SNP most linked to the *SOCS2* SNP (0.229 Mb) was the interval [0.22, 0.24], for which we obtained a mean r^2 of 0.049 on the whole chip. The *SOCS2* SNP mutation was therefore in strong LD with some of the other SNPs in the region.

Genetic parameters

The (co)variance parameters estimated and used in this study are presented in Additional file 1: Figure S3. The variance estimates for the single-trait models were very similar, whether estimated from pedigree or genomic relationships. Heritabilities were 0.50 and 0.61–0.62 for FC and PC, respectively. For Milk Yield (MY), Fat Yield (FY) and Protein Yield (PY), they were 0.37, 0.37, and 0.39, respectively. For Teat Angle (TA), Udder Cleft (UC) and Udder Depth (UD), they were 0.39, 0.34, 0.27–0.28 respectively, and for Lactation Somatic Cell Score (LSCS) 0.17–0.18. Similar results were obtained using the two-trait models (Additional file 1: Figure S3).

Genetic correlations between the *SOCS2* gene content trait and the other traits, and the genetic variances explained

by the *SOCS2* gene using the pedigree-based Gene Content method are presented in Table 1. The absolute values of the genetic correlations ranged from 0.02 (TA) to 0.34 (LSCS). The six traits most correlated with the *SOCS2* gene content trait were: LSCS ($r_g = 0.34$), PY ($r_g = 0.29$), MY ($r_g = 0.25$), UD ($r_g = -0.19$), FY ($r_g = 0.18$) and FC ($r_g = -0.14$). This was confirmed by the genetic variances explained by the *SOCS2* gene that ranged from 0.05% (TA) to 11.24% (LSCS).

GBLUP and WssGBLUP methods improve prediction accuracies

The prediction accuracies and gains obtained with the different evaluation methods and traits are shown in Table 2. Prediction accuracies ranged from 0.498 to 0.561 for MY, from 0.330 to 0.486 for FY and PY, and from 0.684 to 0.762 for FC and PC. They ranged from 0.421 to 0.471 for LSCS, and from 0.336 to 0.538 for udder type traits.

Table 1 Genetic correlations between the *SOCS2* gene content trait and the traits of interest (r_g) and genetic variances (σ_g^2) explained by the *SOCS2* gene obtained using the pedigree-based Gene Content method

Trait	r_g with <i>SOCS2</i>	σ_g^2 explained by <i>SOCS2</i>
MY	0.25	6.18%
FY	0.18	3.22%
PY	0.29	8.55%
FC	-0.14	1.88%
PC	-0.06	0.41%
LSCS	0.34	11.24%
TA	-0.02	0.05%
UC	-0.07	0.56%
UD	-0.19	3.71%

Abbreviations: MY Milk Yield, FY Fat Yield, PY Protein Yield, FC Fat Content, PC Protein Content, LSCS Somatic Cell Score, TA Teat Angle, UC Udder Cleft, UD Udder Depth

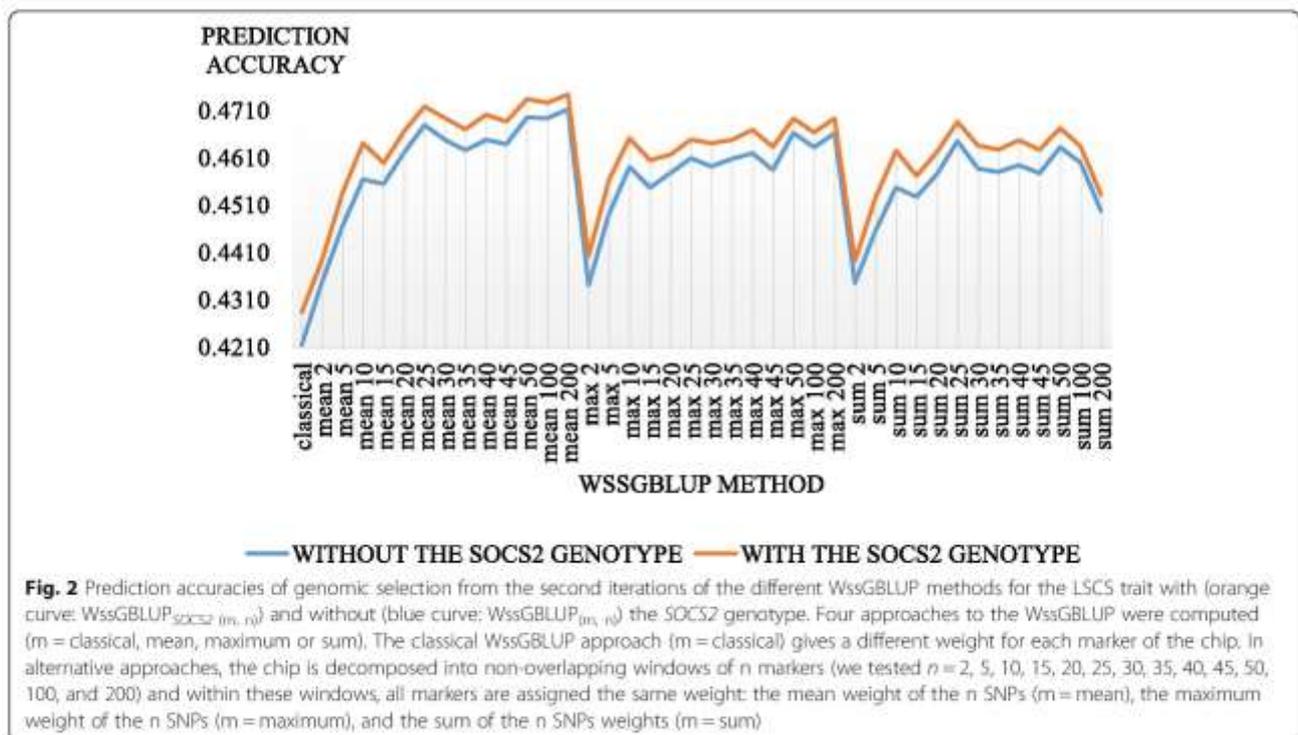
Table 2 Prediction accuracies of different genetic evaluation methods for each trait using information about the *SOCS2* gene or not

		Trait								
		MY	FY	PY	FC	PC	LSCS	TA	UC	UD
Prediction accuracy using	Pedigree-based BLUP	0.507	0.389	0.330	0.693	0.684	0.421	0.451	0.477	0.336
	ssGBLUP	0.549	0.450	0.463	0.724	0.745	0.454	0.523	0.473	0.423
	ssGBLUP _{SOCS2}	0.550	0.450	0.465	0.724	0.745	0.456	0.523	0.473	0.424
	WssGBLUP _(classical, 1)	0.498	0.422	0.437	0.723	0.730	0.421	0.538	0.473	0.452
	The best WssGBLUP _(m, n) method	0.561	0.461	0.486	0.739	0.762	0.471	0.538	0.504	0.460
Pedigree-based Gene Content		0.557	0.430	0.405	0.698	0.688	0.438	0.448	0.512	0.366
Gain in prediction accuracy between	Pedigree-based BLUP & ssGBLUP	8.25%	15.54%	40.34%	4.41%	8.95%	7.80%	15.84%	-0.96%	26.07%
	ssGBLUP & the best WssGBLUP method	2.16%	2.46%	5.04%	2.06%	2.32%	3.77%	2.80%	6.59%	8.75%
	Without & with the <i>SOCS2</i> SNP among the markers (average within the WssGBLUP _(m, n) methods)	0.22%	0.13%	0.51%	0.02%	0.10%	1.06%	-0.02%	0.02%	0.26%
	Pedigree-based BLUP & pedigree-based Gene Content	8.88%	9.54%	18.64%	0.68%	0.57%	3.80%	-0.89%	6.72%	8.33%
Parameters of the best WssGBLUP _(m, n) method		Maximum 100	Maximum 200	Maximum 200	Maximum 40	Maximum 45	Mean 200	Classical	Sum 30	Maximum 5

Abbreviations: MY Milk Yield, FY Fat Yield, PY Protein Yield, FC Fat Content, PC Protein Content, LSCS Somatic Cell Score, TA Teat Angle, UC Udder Cleft, UD Udder Depth

In our study, the highest accuracies were obtained by using alternative WssGBLUP approaches. Figure 2 indicates the prediction accuracies for the LSCS trait (as example) with the various WssGBLUP methods. The average gain in accuracy between WssGBLUP methods with and without the *SOCS2* genotype was +1.06% and the best accuracy (0.471) was obtained with the WssGBLUP_(Mean, 200) method.

The gain in prediction accuracy from pedigree-based BLUP to ssGBLUP, the currently used genomic method, was on average +14.03% (Table 2). An average gain in prediction accuracy of +3.99% was obtained from ssGBLUP to the best WssGBLUP method (for each trait independently). The average gain in prediction accuracy between all the WssGBLUP methods and



WssGBLUP_{SOCS2} was +0.26%. The highest gain (+1.06%) was obtained for LSCS.

Genetic trends of EBVs relative to SOCS2 and polygenic components using the Gene Content method

The average gain in prediction accuracy between pedigree-based BLUP and pedigree-based Gene Content was +6.25%. This gain represents the improvement in prediction accuracy for a trait when genome-wide data (54 K herein) are not available and information about a point mutation is known and is included in a multi-trait genetic evaluation model.

The Gene Content method was used to obtain EBVs for the polygenic component (EBV_{polygen}), excluding the effect of the SOCS2 gene, EBVs associated with the SOCS2 gene (EBV_{SOCS2}), as well as estimated breeding values for the trait of interest (EBV_{trait}), for each trait. The EBV_{polygen} and EBV_{trait} values were very similar for all the traits (Spearman correlation of 0.99), except for LSCS (Spearman correlation of 0.87), the trait most associated with SOCS2. The genetic trends of both the polygenic (EBV_{polygen}) and gene (EBV_{SOCS2}) components of the LSCS trait over the years are provided in Fig. 3. This graph shows a strong decrease in the EBVs for LSCS since 2004. This decrease has been due partly to the reduced effect of the SOCS2 gene mutation on the trait (decreased frequency of the deleterious allele), but also to the reduction (improvement) of the polygenic component determining the trait.

SNP effects estimated by using WssGBLUP methods

The estimated SNPs effects and percentages of the explained variance were determined by applying WssGBLUP methods, with and without the SOCS2 genotype among the markers. The QTL regions (positions on ovine genome assembly v4.0) found by applying the best alternative WssGBLUP method for each trait (SOCS2 SNP included in the markers), based on a threshold of 1% of genetic variance explained, are presented in Table 3. According to the SNPs effects (Additional file 1: Figure S4) and the explained variances (Additional file 1: Figure S5), the QTL in the SOCS2 gene region (Table 3) was confirmed for LSCS, with 20 adjacent SNPs (including the SOCS2 SNP) explaining as much as 12.00% of the genetic variance. Moreover, this region also influenced PY (4.91% of the variance explained), UD (4.02%), MY (3.94%), FC (2.57%), and UC (1.84%). In addition, among the surrounding SNPs, the SOCS2 SNP exhibited the strongest (or second strongest) effect (Additional file 1: Figure S4).

Several other QTLs were also detected in this study (Table 3). Some of them were trait-specific, such as QTLs on OAR 3 (140.1–141.5 Mb) and 6 (84.7–85.8 Mb), associated with PC, on OAR 19 (44.5–45.6 Mb) and 23 (32.4–33.9 Mb), associated with UC, and on OAR 13 (63.4–64.5 Mb) and 20 (48.8–49.8 Mb), associated with FC and LSCS, respectively. The other three QTLs seemed to be associated with several traits, such

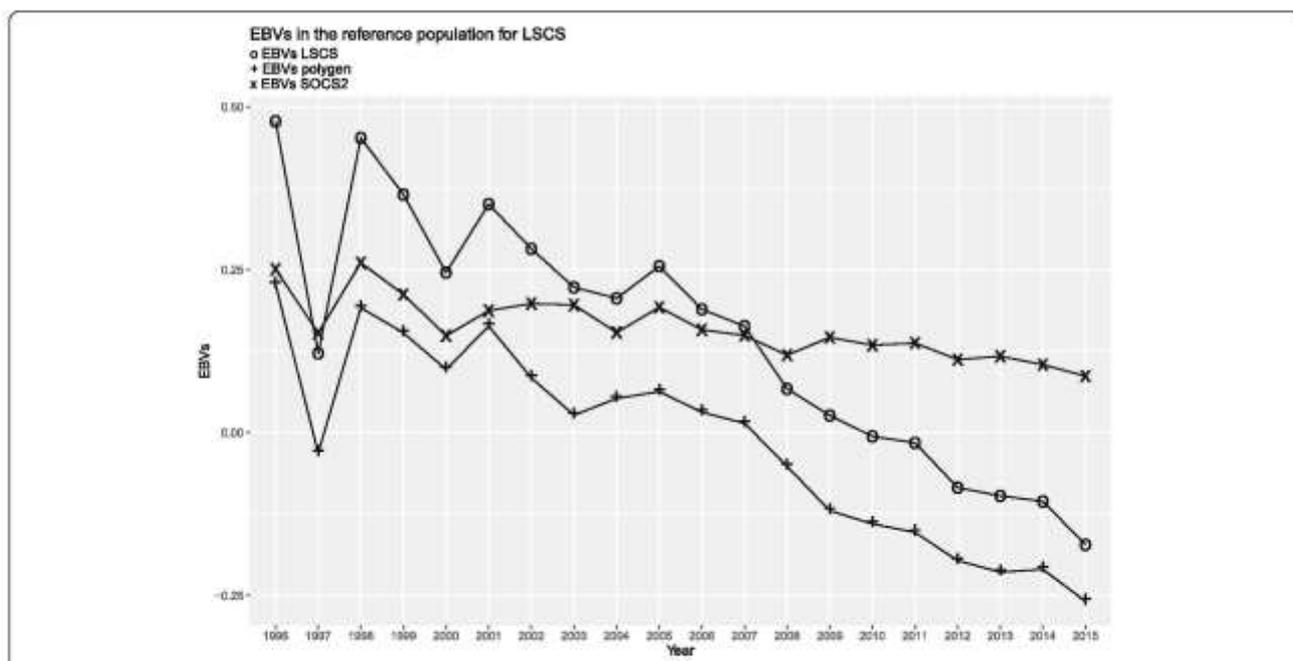


Fig. 3 Genetic trends over the years in the reference population (5343 rams with reference performances, i.e., daughter yield deviations, born between 1996 and 2015), of the EBVs for LSCS using the Gene Content method which enables the polygenic component (EBV_{polygen}), excluding the effect of the SOCS2 gene, and the breeding value due to the gene effect (EBV_{SOCS2}), to be distinguished

Table 3 QTL (Quantitative Trait *Loc*) regions (positions on ovine genome assembly v4.0) found using the best alternative WssGBLUP method for each trait (SOCS2 SNP included among the markers), and based on a threshold of 1% of genetic variance explained

OAR	QTL region (Mb)	Trait associated with the QTL	Genetic variance explained in the trait-specific QTL region ^a (%)	Trait-specific QTL region ^a (Mb)
3	128.3 - 130.5	LSCS	12.00	129.1 - 130.5
		PY	4.91	129.0-130.3
		UD	4.02	129.1-130.5
		MY	3.94	128.3-129.6
		FC	2.57	129.1-130.5
		UC	1.84	129.1-130.5
	136.3 - 137.6	PC	4.61	136.3-137.3
		FC	1.20	136.4-137.6
	140.1 - 141.5	PC	1.24	140.1-141.5
	6	84.7 - 85.8	PC	5.95
11	33.1 - 34.9	FC	6.65	33.4-34.9
		PY	2.25	33.3-34.5
	LSCS	1.98	33.3-34.5	
	MY	1.31	33.1-34.3	
13	63.4 - 64.5	FC	1.17	63.4-64.5
17	8.5 - 10.5	MY	1.14	8.5-9.7
		FC	1.09	8.6-9.9
		PC	1.02	9.0-10.5
19	44.5 - 45.6	UC	2.33	44.5-45.6
20	48.8 - 49.8	LSCS	3.09	48.8-49.8
23	32.4 - 33.9	UC	2.01	32.4-33.9

Abbreviations: OAR Ovis Aries, QTL Quantitative Trait Loci, Mb Megabase, MY Milk Yield, FY Fat Yield, PY Protein Yield, FC Fat Content, PC Protein Content, LSCS Somatic Cell Score, TA Teat Angle, UC Udder Cleft, UD Udder Depth
^aTrait-specific QTL regions are regions where 20 adjacent SNPs explain the highest value of genetic variance of the trait

as QTLs on OAR 3 (136.3–137.6 Mb), associated with PC and FC, on OAR 11 (33.1–34.9 Mb), associated with MY, PY, LSCS and FC, and on OAR 17 (8.5–10.5 Mb), associated with MY, FC and PC.

Discussion

Genetic parameters estimation

The genetic parameters estimation with pedigree or genomic relationships gave similar results for the one-trait methods. In this study the obtained heritabilities were higher for all the traits than previously found in the Lacaune breed. The heritability for LSCS ($h^2 = 0.17 - 0.18$ in our study) was previously estimated at between 0.12 and 0.15 by Barillet et al. (2001), Rupp et al. (2003) and Barillet et al. (2007) [20–22]. These authors also reported lower heritabilities for MY, FC and PC (0.28–0.34, 0.41–0.50 and 0.51–0.63, respectively). Similar

lower results were found for FY and PY (0.26 and 0.28, respectively) [22], and also for the three udder-type traits (0.33–0.35, 0.26–0.32 and 0.19–0.26, for TA, UC and UD, respectively) [22, 23]. These discrepancies could be due, at least in part, to the model as previous studies were based on sire models whereas we used animal models. Other explanations include increased genetic variance within the population (good management of matings in farms, for example) and/or decreased environmental variance (possibly due to the homogenization of breeding practices, for example).

Detection of QTLs and quantification of their effect using genomic evaluation methods

Our study shows that genomic evaluation methods, which involve weighted approaches and thus an initial step to estimate the effects of SNPs on different traits, can be applied to detect and confirm QTLs [13, 14]. This approach is preferable to a single SNP GWAS (Genome-Wide Association Study) approach because the phenotypes of ungenotyped animals can be directly incorporated, without computing pseudodata, as suggested by Wang et al. (2012) [13].

We initially validated, as expected, the QTL for LSCS on chromosome 3 associated with the *SOCS2* gene point mutation first discovered in Lacaune sheep by Rupp et al. (2015) [19]. When we applied WssGBLUP approaches, this region was found to explain 12% of the genetic variance, as already reported by Rupp et al. (2015) [19]. We also confirmed the pleiotropic effect of this region and its association with milk production traits and UD and were able to quantify its effects on these traits, by applying WssGBLUP approaches. Indeed, this region was found to explain 6.2% of the genetic variance for MY, compared to the 4.4% estimated by Rupp et al. (2015) [19]. The association of this *locus* with UD found in our study might be explained by an indirect effect of individual body size.

We then discovered another pleiotropic QTL on chromosome 11 (33.1–34.9 Mb). This QTL was associated with milk production traits and LSCS, with 1.31, 2.25, 6.65 and 1.98% of the genetic variance explained for MY, PY, FC and LSCS, respectively. This region had previously been associated with LSCS in Lacaune sheep [19] (35.8–41.3 Mb, ovine genome assembly v3.1). A similar pleiotropic QTL was found very near to the orthologous region on caprine chromosome 19 in Saanen goats. Indeed, Martin et al. (2018) [24] reported a pleiotropic QTL (chromosome 19: 24.5–26.9 Mb, caprine genome assembly CHIR_1.0) for milk production and udder traits including MY, FY, PY, udder floor position, and rear udder attachment. This QTL was validated by Teissier et al. (2019) [17] (top 10 SNPs with the highest SNP weights on chromosome 19 located between 26 and 28 Mb, caprine genome assembly CHIR_1.0). It was then confirmed by Oget et al. (2018) [25] (chromosome 19: 22.8–28.9 Mb, caprine genome assembly

ARS1) who found that this QTL was also associated with the lifespan of livestock and semen production. However, due to the large number of genes present in the region (47 protein-coding genes, NCBI ovine genome assembly v4.0), proposing suitable candidate genes remains difficult.

In addition to the QTLs associated with LSCS on chromosomes 3 and 11, we also found a QTL on chromosome 20 (48.8–49.8 Mb) that explained 3.1% of the genetic variance. This QTL had previously been detected in sheep [19] (48.6–48.8 Mb, ovine genome assembly v3.1). Sixteen protein-coding candidate genes in this region, two of them related to immune defense (*SERPINB1*, *RIPK1*), were annotated. These three QTLs, associated with LSCS in our study, accounted for as much as 17% of the genetic variance.

Regarding PC, a QTL explaining 5.95% of the genetic variance of this trait was detected on chromosome 6 in a narrow region (84.7–85.8 Mb) known as the casein gene cluster. This region encodes for the caseins (*CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*; 85.0–85.2 Mb), which are the main proteins in milk. Caseins are responsible for milk coagulation, a fundamental step in the preparation of cheese from raw milk. Previous association studies based on microsatellite markers had already highlighted an association of the ovine chromosome 6 with PC, suggesting the role of casein genes, but the confidence intervals obtained with those low-density marker panels were very large [26].

Weighting SNPs improves genomic evaluation by capturing QTL regions

In 2015, genomic selection was implemented in Lacaune dairy sheep following two comparative studies of evaluation accuracy involving different approaches. Duchemin et al. (2012) [10] compared BLUP, Bayes C π , Partial Least Squares (PLS), and sparse PLS methods and reported that including markers in the models increased EBV accuracies by 18 to 25%, depending on the trait (MY, FC et SCS), with minor differences between the genomic approaches. Based on these results, Baloché et al. (2014) [11] adopted BLUP-like methods to implement a single-step model in the evaluation. These authors compared three strategies: pseudo-BLUP (using all rams and DYDs), pseudo-ssGBLUP (using all rams and DYDs), and regular ssGBLUP (using all phenotypes and pedigree in an animal model) and obtained the best results with regular ssGBLUP. Based on these results, the ssGBLUP method is now used for the routine evaluation of Lacaune dairy sheep [11], and hence was the reference method adopted in our study. Using the same ssGBLUP method, we obtained better evaluation accuracies than those of the two previous studies for production traits: +14.4%, +1.9% and +6.1% for MY, FC and PC, respectively. One explanation for this result could be the larger size of the population of genotyped individuals. Indeed, 2892 individuals were genotyped in Baloché et al. (2014) compared to 9844 genotyped animals in our study. However, we

obtained lower accuracies for LSCS and the three udder-type traits than Baloché et al. (2014) [11]. No straightforward explanation has been found for this surprising result.

Weighting alternative strategies in the evaluation models was found to provide more accurate results than ssGBLUP for all the traits in our study, with an average gain of +3.99%, even when no QTL was detected for the trait (teat angle, for example). These results are in slight disagreement with those obtained in goat [17] where the addition of a weighting strategy increased accuracies only for traits that exhibited QTLs. Indeed, the large pleiotropic QTL on chromosome 19 in the Saanen breed [17] allowed an increase in accuracy while in the Alpine breed, with no QTL segregating for most of the traits, WssGBLUP did not provide any significant gain. This disagreement might be explained by the fact that we retained the best alternative method for each trait and did not use the same strategy for all traits, as was done in goats. Indeed, Teissier et al. (2019) [17] used a window size of 40 SNPs for all traits. In our study, depending on the trait, the best accuracies were obtained by using an alternative WssGBLUP strategy with a large window size (100–200 SNPs) for MY, FY and PY, and for LSCS, a medium window size (40–45 SNPs), for FC and PC and a small window size (1–30 SNPs), for the udder-type traits (TA, UC, and UD). Due to the differences in genetic determinism of each trait, (confirmed by the estimated effects of the SNPs in this study, Additional file 1: Figure S4), and possible variation in QTL size, it seemed more appropriate to consider an optimal evaluation strategies for each trait individually. With fine-tuning for each trait, WssGBLUP always proved better than non-weighted ssGBLUP. This highlights a means of increasing genetic progress by taking QTLs into account during genomic evaluation. A further advantage of this method is that no identification of the QTLs, in a dedicated preliminary study, is required.

Accounting for causal mutations in genomic evaluations

As a case study, we tested different strategies to include the effect of a known causal mutation (*SOCS2* gene) in genomic evaluations. The gain between methods, with and without the mutation in the chip, was limited: an average gain of +0.26% and up to +1.06% for LSCS, the trait most influenced by the mutation. This result suggests that the strong LD between the mutation and the surrounding SNPs (0.17) was sufficient to allow accurate estimation of the genomic breeding values without the genotype at the causal mutation, as stated in Goddard (2009) [27]. Including causal SNPs might be of greater interest in the case of lower SNP density and LD around the QTLs or in the case of a stronger effect of the causal mutation.

By applying the Gene Content method, the genetic value (EBV) for a trait can be separated into a genetic value resulting from the effect of a given gene, and a polygenic

component resulting from all the other QTLs. In this study, the frequency of the *SOCS2* SNP allele was found to decrease for AI rams (from 0.21 in 2005 to 0.09 in 2017), which might be explained by the introduction of the SCS trait, considered as a proxy for mastitis, into the breeding objectives in Lacaune dairy sheep in 2005 [28, 29]. Indeed, the current relative weight for SCS is 25% in the total merit index. In addition, breeding companies are advised not to retain individuals that are homozygous for the mutation, due to its very negative effect on health. In our study, the slope (Fig. 3) of the decrease over time in the breeding value due to the *SOCS2* gene effect is less pronounced than that due to the polygenic effect. It can therefore be deduced that decreasing the allele frequency of the *SOCS2* mutation over time in the population (unfavourable for the SCS trait) (Fig. 1) does not prevent a favourable trend in other mastitis resistance genes. Thus, in Lacaune sheep, applying a weighting strategy to markers that are strongly associated with the trait of interest (SCS), as with the WssGBLUP approaches, will improve accuracy and therefore response to selection, and may be sufficient to increase mastitis resistance in the whole population.

Two other alternatives to the methods tested in this study would be (i) to include the causal mutation (or major gene) in the evaluation model as a fixed effect [30] and (ii) to use a mixed inheritance model [31]. These alternatives would require knowledge of the gene of interest for a large majority of individuals with a given phenotype, which was not the case in our study. Indeed, the single-step approach makes it possible to start from raw phenotypes and therefore from a very large number of observations (3,575,614 lactations for MY for example). Considering that *SOCS2* information was available for only 9844 individuals (including 1517 females with phenotypes at best), the number of missing data was too large to apply this method to our study.

The presence of a major gene raises two questions: (1) is the polygenic breeding value overestimated when the major gene is ignored [32], (2) what is the risk of reducing total genetic variance over generations of selection [33]. The Gene Content method offers a solution to both questions because it allows the genetic component due to the major gene to be separated from the remaining polygenic component. It might then be possible to manage these two values separately according to the selection objectives i.e., eradicate or fix a major gene allele while maintaining or increasing the polygenic component associated with the trait. The Gene Content method is promising because it allows the genetic variability, i.e. other QTLs or regions with low effects on the same trait, to be preserved in the population, whether the major gene has a stronger effect than the polygenic component for trait prediction or not. This method is also of interest if the gene has a pleiotropic effect on both selected and non-selected traits e.g., a mutation

with a favourable effect on a production trait but associated with defects or disease not included in the breeding scheme.

Conclusions

This study highlights the interest of weighted alternative methods (WssGBLUP) for capturing QTL and major genes in genetic evaluation models. These alternatives increase the accuracy of the predicted genetic values and therefore the expected genetic gains in the population. On the other hand, another approach (Gene Content) tested in this study showed promise for the genetic management of particular traits since it allows the genetic component due to a major gene, to be dissociated from the remaining polygenic component. This latter method is also interesting for populations that have not been genotyped with SNP chips but for which information about a major gene is available. The results of this study pave the way for an improved management of trait genetics, directly applicable to the selection schemes in different livestock sectors.

Methods

Animals and phenotypes

The performances of Lacaune sheep registered in the French official milk recording scheme since 1960 and available from the national database (Centre de Traitement de l'Information Génétique, CTIG, Jouy-en-Josas, France) were used for this investigation. The corresponding pedigree information was obtained from the official livestock data system (Ministerial Order NOR: AGRT1431011A, 24th March 2015, Ministry of Agriculture, France).

Nine traits, included in routine genetic evaluations [28, 29, 34], were considered: milk, fat and protein yields, fat and protein contents, somatic cell score (SCS) and three udder-type traits: teat angle, udder cleft, and depth.

The first three lactations were retained for traits related to milk production and SCS. Briefly, milk yield was measured monthly. Milk yield per lactation (MY) was estimated using the Fleischmann method and adjusted for milking length over a reference period of 220 days [35]. SCC, fat and protein contents were measured three times per lactation on average. Lactation traits for fat (FC) and protein (PC) contents were defined as the weighted mean of test days adjusted for milk. Weights were defined according to lactation length and parity. Lactation traits for fat (FY) and protein (PY) yields were the product of MY and corresponding FC and PC. Test-day SCC were log-transformed to somatic cell score ($SCS = \log_2(SCC/100) + 3$) [36] to normalize the data distribution and were averaged per lactation to compute the analyzed trait LSCS, as described in Rupp et al. (2003) [21].

During the first lactation, three udder-type traits including teat angle (TA), udder cleft (UC) and udder

Table 4 Description of the Lacaune dairy sheep dataset used for genetic evaluation

Trait	Mean \pm SD	Number of			
		Ewes	Lactations ^a	Individuals in the pedigree file	Rams in the validation population ^b
MY (L)	292.91 \pm 85.46	1,503,148	3,575,614	1,651,901	264
FY (g)	213.14 \pm 51.56	1,124,636	1,841,351	1,336,060	263
PY (g)	169.72 \pm 39.33				
FC (g/L)	66.54 \pm 8.51				
PC (g/L)	52.89 \pm 4.63				
LSCS	3.12 \pm 1.56	769,929	1,321,411	1,031,375	263
TA	7.15 \pm 1.06	349,134	349,134	349,134	250
UC	5.04 \pm 1.26	349,132	349,132	653,908	249
UD	6.42 \pm 0.70	349,132	349,132	653,907	253

Abbreviations MY Milk Yield, FY Fat Yield, PY Protein Yield, FC Fat Content, PC Protein Content, LSCS Somatic Cell Score, TA Teat Angle, UC Udder Cleft, UD Udder Depth

^aMY, FY, PY, FC, PC and LSCS were measured during the first three lactations (lactation average); TA, UC, and UD were measured once during the first lactation

^bRams born in 2015 with Estimated Breeding Values (EBVs) and reference performances, i.e. Daughter Yield Deviation (DYD)

depth (UD) were scored over a linear range of 1 to 9, as described in Marie-Etancelin et al. [23].

Performances were available for ewes since birth years 1978 (MY), 1987 (milk composition traits), 1999 (LSCS) and 2000 (udder-type traits). Data were included up to the birth year 2015. Descriptive statistics for the performance and pedigree files of each trait are given in Table 4.

Genome-wide genotyping data

Genotyping data used in genomic evaluations in this study were available for 9844 individuals that had been genotyped with the medium-density Illumina Ovine 54K SNPs chip [37] for the current genomic selection program [12], or as part of former research projects: “SheepSNPQTL”, “Sustainable Solutions for Small Ruminants”, “Roquefort’in”, and “PhénoFinLait”. This genotyping data was derived from Artificial Insemination (AI) rams ($N = 8327$), born between 1996 and 2015, and progeny-tested by the two Lacaune breeding companies (OVI-TEST -La Glène, Saint-Léons, France- and Confédération Générale de Roquefort -Le Bourguet, Vabres l’Abbaye, France-), and from ewes ($N = 1517$), born between 2004 and 2013, and used for QTL detection programs.

DNA extraction from blood samples and genotyping were performed at the Laboratoire d’Analyses Génétiques pour les Espèces Animales (LABOGENA -Jouy en Josas, France-; www.labogena.fr). SNPs were remapped on version 4.0 of the ovine genetic map (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2/). Quality control was performed as part of the routine pipeline for Lacaune genotyped animals, as described in Baloché et al. [11] with slight modifications. In brief, SNPs with a MAF lower than 1%, Hardy Weinberg disequilibrium ($P < 10^{-5}$) and a call rate lower than 97% were removed. After edits, 37,941 out of 54,241 SNPs remained for the analyses.

SOCS2 genotypes and imputation of missing data

Genotypes for the point mutation of interest in the *SOCS2* gene (hereafter called *SOCS2* SNP, rs868996547, *Ovis aries* -OAR- chromosome 3, position 129,557,942 on ovine genome assembly v4.0) were available for 4297 animals. The data were derived from two datasets. First, KASPar™ tests (described in Rupp et al. [19]) were obtained as part of the “Sustainable Solutions for Small Ruminants” and “REIDSOCS” (ANR-16-CE20-0010 funded by the ANR -Paris, France) projects for 1413 AI rams and 248 ewes from one INRA (Institut National de la Recherche Agronomique) experimental farm (La Fage -Roquefort-Sur-Soulzon, France-) born during 2002–2003. These 1661 individuals are included in the 9844 individuals genotyped with the 54K SNPs chip presented in the previous section. Second, young rams ($N = 2636$) which entered breeding centers in 2017 were low-density genotyped with the International Sheep Genomics Consortium (ISGC) panel which includes 1500 SNPs [38]. This chip also contains the *SOCS2* mutation as a SNP and suitable genotypes were subsequently extracted. In addition, these rams were imputed from low to 54K density as part of the routine genomic evaluation [39]. All 4297 animals were then genotyped both for *SOCS2* and for the 54K SNPs panel.

The *SOCS2* genotype was then imputed using the FImpute v2.2 software [40] for the 9844 individuals genotyped with the 54K SNPs chip. For this imputation step, individuals from AI centers born in 2016 and genotyped with the 54K SNPs chip were also added to the data set to fill the missing year that connected young males born in 2017 to the rest of the genotyped population. Thus 10,432 out of the total of 14,729 individuals genotyped with the 54K SNPs chip had missing *SOCS2* locus information and required imputation. A cross-

validation test was performed to assess the accuracy of imputation by designing an appropriate validation population for which the *SOCS2* genotypes were ascribed to missing values. The validation population represented one-third of the total number of *SOCS2* genotypes, i.e. 1432 individuals selected at random. The accuracy of imputation was calculated by counting the errors between the true and imputed genotypes in the validation population. The accuracy of the imputation corresponded to the concordance rate (CR), with $CR = 1 - error_rate$, where $error_rate = \frac{\text{number of errors}}{2 \times \text{number of imputed individuals}}$

Linkage disequilibrium (LD)

After imputing the *SOCS2* genotypes, we then computed the linkage disequilibrium (LD) in the surrounding chromosomal region. Using the Haploview v4.2 software [41], the square correlation coefficient r^2 measure of LD [42] was calculated between each pair of SNPs for *SOCS2* and the 40 closest markers (Additional file 1: Figure S2).

The LD of the *SOCS2* region was then compared with the mean LD of the chip by computing the r^2 between all SNP pairs in 10-Megabase (Mb) windows within the chromosomes. The r^2 were grouped into categories based on the distance between SNP (every 0.02 Mb) (Additional file 1: Figure S1).

Genomic prediction methods

EBVs were computed for all animals in the pedigree files (Table 4) using the following evaluation approaches. With the first set of methods, a single-trait approach was used for all nine traits of interest. With the second set of methods, a multi-trait approach called Gene Content, developed by Legarra and Vitezica [15], was used in which the *SOCS2* genotype was considered as a trait in bivariate evaluations with MY, FC, PC, FY, PY, LSCS, TA, UC and UD. All methods were based on a single-step process [5],

e.g. use of all data from females together with the pedigree, and genomic information if available. These different approaches are summarized in Table 5.

Single-trait approaches

For single-trait approaches, we applied the following model (1) to the five milk production traits (MY, FC, PC, FY, and FC) and LSCS:

$$y = X\beta + Zg + Wp + \epsilon \tag{1}$$

where y is the observation vector for the trait (female lactation performances) and β is a vector of fixed effects. The fixed effects for each trait are listed in Table 6. g is a vector of random additive genetic effects assumed to be normally distributed $N(0, H\sigma_g^2)$, with H the relationship matrix. p is a vector of random permanent environmental effects assumed to be normally distributed $N(0, I\sigma_p^2)$ and ϵ is a vector of random residuals that is normally distributed $N(0, I\sigma_e^2)$. X is the incidence matrix relating phenotypes to the fixed effects (β), Z is the design matrix allocating phenotypes to breeding values (g) and W is the incidence matrix relating phenotypes to permanent environmental effects (p).

Since the three udder-type traits had only one record per female, we removed the random permanent environmental term from the eq. (1) and applied the following model with the same parameters as in (1):

$$y = X\beta + Zg + \epsilon \tag{2}$$

In the pedigree files, we added 24 unknown parent groups defined as follows: animals born before 1960, cohorts born within 10-year windows up to 2000, cohorts born within 5-year windows up to 2010, cohorts born within 2-year windows up to 2014, and finally animals born in 2015.

Table 5 Description of the different genetic evaluation models based on a single-step approach and using information about the *SOCS2* gene or not

Approach	Model	Use of <i>SOCS2</i> data	Information used in the relationship matrix		
			Pedigree	54K SNPs	<i>SOCS2</i> SNP
Single-trait	Pedigree-based BLUP	No	Yes	No	No
	ssGBLUP	No	Yes	Yes	No
	ssGBLUP _{<i>SOCS2</i>} ^a	Yes	Yes	Yes	Yes
	WssGBLUP _(m, n) ^b	No	Yes	Yes	No
	WssGBLUP _{<i>SOCS2</i> (m, n)} ^b	Yes	Yes	Yes	Yes
Multiple-trait	Pedigree-based Gene Content	Yes (as a trait)	Yes	No	No

Abbreviations: GBLUP Genomic Best Linear Unbiased Prediction, ss single-step, W Weighted

^aThe term *SOCS2* here means that the *SOCS2* SNP has been added to the 54K SNPs of the chip

^bFour approaches to the WssGBLUP were computed (m = classical, mean, maximum or sum). The classical WssGBLUP approach (m = classical) gives a different weight for each marker of the chip. In alternative approaches, the chip is decomposed into non-overlapping windows of n markers (we tested n = 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, and 200) and within these windows, all markers are assigned the same weight; the mean weight of the n SNPs (m = mean), the maximum weight of the n SNPs (m = maximum), and the sum of the n SNPs weights (m = sum)

Table 6 Description of fixed effects for the evaluation models of each phenotype

Trait	Fixed effects	Number of levels
MY	• herd within year and within parity	42259
	• age at delivery within year and within parity	514
	• month at delivery within year and within parity	702
	• time between delivery and first OMR within year and within parity	585
FY	• herd within year and within parity	20783
	• age at delivery within year and within parity	269
PY	• quality control within year and within parity	348
FC		
PC		
LSCS	• herd within year and within parity	14306
	• age at delivery within year and within parity	206
	• month at delivery within year and within parity	312
TA	• herd within year	3672
UC	• interaction between examiner and the time difference between milking and scoring, within herd	2346
	• interaction between age at delivery and lactation stage, within year	160
UD	• number of lambs within year	30

Abbreviations: MY Milk Yield, FY Fat Yield, PY Protein Yield, FC Fat Content, PC Protein Content, LSCS Somatic Cell Score, TA Teat Angle, UC Udder Cleft, UD Udder Depth

We modeled the relationship matrix H using the different approaches summarized in Table 5. Briefly, only pedigree information was used for the pedigree-based BLUP and a combination of pedigree and genomic information for the ssGBLUP method, as described in Legarra et al. [5].

Next, four WssGBLUP methods were applied. These differed from the ssGBLUP method in that the genetic relationship matrix was altered by weighting the chip markers, the weights being iteratively derived from the decomposition of EBVs into marker effects. Indeed, SNP effects can be deduced from EBVs in the genomic-based single-trait approaches (eqs. (1) and (2)), as modeled in Wang et al. (2012) [13]: $\hat{a} = DM' [MDM]^{-1} \hat{g}_{gen}$. In this equation, \hat{a} is a vector of estimated SNP effects, D is a diagonal matrix of weights (set at 1 in the ssGBLUP method), M is the centered matrix of SNP genotypes, and \hat{g}_{gen} is the vector of EBVs from genotyped animals only.

Wang et al. [13] showed that WssGBLUP was sufficient, with only very few iterations, to attain a maximum accuracy of EBV. Similarly to Wang et al. (2012), and validated in Teissier et al. [16, 17], the highest prediction accuracies in our study were obtained after two iterations (results not shown) and therefore only the results after two iterations are provided. The decrease in prediction accuracy after the second iteration in the WssGBLUP approaches could be due to excessive weighting of SNPs associated with a few high effect QTLs, and reduced weighting of numerous low-effect QTLs.

Alternative weighting methods, that assign the same weight to several adjacent chip markers, have been proposed by Zhang et al. [14]. In our study, we computed three alternative methods using WssGBLUP: (1) the mean weight of n SNPs (with n the number of adjacent SNPs with non-overlapping windows), (2) the maximum weight of n SNPs, and (3) the sum of n SNPs weights. Weights were calculated as described in Teissier et al. [16]. Briefly, calculations of the weights used in the diagonal matrix D in these alternative methods were based on the variances of SNPs effects estimated with the first step ssGBLUP. After assigning the same value (mean, sum or maximum) to the n markers of a window, the vector of marker weights was then normalized so that the sum of all weights remained constant and equal to the total number of SNPs. Several window sizes with varying number of SNPs were used ($n = 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, \text{ and } 200$). Hereafter, these methods are designated WssGBLUP_(m, n), where m is the method used to calculate the weights and n the number of adjacent SNPs with non-overlapping windows.

Estimates of SNP effects can also be used to estimate the genetic variance of the trait explained by each SNP i effect: $\hat{\sigma}_{a,i}^2 = \hat{a}_i^2 2p_i(1-p_i)$, where p_i is the allele frequency of SNP i . The explained variances will be described for 20 adjacent SNPs in the result and discussion sections.

Multiple-trait gene content approaches

In the Gene Content method [15] EBVs are estimated for a given trait by simultaneously considering information about a given genotype as a second trait (hereafter called the *SOCS2* gene content trait) in a two-trait approach. Accordingly, the following model (3) was applied to the five milk production traits and LSCS:

$$\begin{cases} y = X\beta + Zg + Wp + \varepsilon \\ y_T = \mu_T + Z_T g_T + \varepsilon_T \end{cases} \quad (3)$$

where y is the observation vector for the trait (female lactation performance) and parameters of the model $\beta, g, p, \varepsilon, X$ and W are the same as in eq. (1). y_T is a vector of the *SOCS2* gene content trait, i.e. the number of copies of the mutant T allele carried by each animal (0, 1 or 2). Missing values were set for ungenotyped individuals. μ_T is the mean fixed effect of the *SOCS2* T allele, Z_T is the incidence matrix relating observations to the random genetic effect (g_T) of the *SOCS2* gene content trait, which was assumed to be normally distributed such that $\hat{\sigma}_{g_T}^2 = H\hat{\sigma}_{g_T}^2$ and $\hat{\sigma}_{g_T}^2 = 2\hat{f}_T(1-\hat{f}_T)$, with f_T the T allele frequency, and ε_T the random residual error.

As before, the model (4) was simplified for the three non-repeated udder-type traits:

$$\begin{cases} y = X\beta + Zg + \varepsilon \\ y_T = \mu_T + Z_T g_T + \varepsilon_T \end{cases} \quad (4)$$

For Eqs. (3) and (4) only the pedigree-based approach was tested, in order to avoid redundant information between the chip and the *SOCS2* gene content trait (copy number of the mutated allele), which meant that the kinship matrix *H* was modeled using only pedigree information (Table 5).

The evaluations were done using BLUP90IOD2 v3.102 software [43].

By applying the Gene Content method, we were able to obtain not only EBVs for the trait of interest (\hat{g}), but also estimates of a breeding value for the polygenic component ($EBV_{polygen}$) that excluded the effect of the *SOCS2* gene, as well as estimates of breeding values associated with the *SOCS2* gene (EBV_{SOCS2}). We calculated these three EBVs as proposed by Legarra and Vitezica (2015) [15]: $EBV_{polygen} = EBV_{trait_of_interest} - EBV_{SOCS2} = \hat{g} - \hat{g}_T \hat{\alpha}$, with $\hat{\alpha} = \frac{cov(g, g_T)}{\hat{\sigma}_{g_T}^2}$, which can be interpreted as the allele substitution effect of the *SOCS2* gene mutation on the trait.

Variance component estimation

Variance and (co)variance (for the Gene Content method) components for models (1), (2), (3) and (4) for each approach and each trait were estimated using a block implementation of Gibbs sampling computed in the GIBBS1F90 v1.44 software [43].

Based on these variance component estimations, the genetic variance explained by the *SOCS2* gene for each trait was calculated by applying the following equation [15]: $explained_variance = \frac{\hat{\alpha}^2 \times \hat{\sigma}_{g_T}^2}{\hat{\sigma}_g^2}$, with α the allele substitution effect of the *SOCS2* gene mutation on the trait described previously.

We also used the estimated variance components to derive the genetic correlation (r_g) between the *SOCS2* gene content trait and the trait of interest with the following equation:

$$\hat{r}_g = \frac{cov(g, g_T)}{\hat{\sigma}_g \times \hat{\sigma}_{g_T}}$$

Prediction accuracy

To validate the EBVs, we added progeny performances from 264 males born in 2015, which had not been used to predict these EBVs, to a validation set (Table 4). We computed the accuracies of the genomic predictions for each model and for each trait using the Pearson correlation between EBVs for the males in the validation

population and DYDs. The numbers of rams in the validation population for each trait with EBVs and DYD are shown in Table 4.

Additional file

Additional file 1: Figure S1 Visualization of linkage disequilibrium ($r^2 \times 100$) between the 40 markers closest to the *SOCS2* point mutation (rs868996547, in the middle). **Figure S2** Visualization of linkage disequilibrium measured as squared correlation coefficient (r^2) according to distance between markers on the 50K ovine SNP chip. **Figure S3** Components estimations according to the different models. One-trait methods correspond to eqs. (1) and (2) and two-traits methods to eqs. (3) and (4). **Figure S4** Manhattan plots of estimated SNP effects using the best WssGBLUP approach for each phenotype (second iteration). On the left are presented analysis without the *SOCS2* genotype among the markers and on the right, with the *SOCS2* genotype (green point). **Figure S5** Manhattan plots of estimated variance explained by 20 adjacent SNPs using the best WssGBLUP approach for each phenotype (second iteration). The horizontal red line represents the threshold of 1% adopted in this study. On the left are presented the analyses without the *SOCS2* genotype among the markers and on the right, with the *SOCS2* genotype. (DOCX 2190 kb)

Abbreviations

AI: Artificial Insemination; ANR: Agence Nationale de la Recherche; BLUP: Best Linear Unbiased Prediction; CR: Concordance Rate; DYD: Daughter Yield Deviation; FC: Fat Content; FEDER: Fonds Européen de Développement Régional; FGE: France Génétique Elevage; FUI: Fonds Unique Interministériel; FY: Fat Yield; GBLUP: Genomic Best Linear Unbiased Prediction; INRA: Institut National de la Recherche Agronomique; ISGC: International Sheep Genomics Consortium; LABOGENA: Laboratoire d'Analyses Génétiques pour les Espèces Animales; LD: Linkage Disequilibrium; MAF: Minor Allele Frequency; MY: Milk Yield; OMR: Official Milk Recording; PC: Protein Content; PY: Protein Yield; QTL: Quantitative Trait Loci; SCS: Somatic Cell Score; SNP: Single Nucleotide Polymorphism; *SOCS2*: Suppressor Of Cytokine Signaling 2; ssGBLUP: Single-step Genomic Best Linear Unbiased Prediction; TA: Teat Angle; UC: Udder Cleft; UD: Udder Depth; WssGBLUP: Weighted single-step Genomic Best Linear Unbiased Prediction

Acknowledgments

The authors are grateful to Andres Legarra, Christèle Robert-Granić, Héliane Larroque and Céline Carillier-Jacquin for their help in interpreting the analyses. They also thank Florent Woloszyn and Julien Sarry for mutation genotyping during the various projects. Claire Oget acknowledges the support of the Agence Nationale de la Recherche (ANR) for her scholarship (Project Reidsocs, ID: ANR-16-CE20-0010).

Authors' contributions

CO adapted the methods for the dataset, performed the analyses, contributed to their interpretation and wrote the draft. MT developed the scripts used for the different genomic evaluation approaches and helped to interpret the analyses. JMA contributed to data collection and performed phenotype calculations. GTK and RR designed the study and helped to interpret the analyses. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the French National Research Agency (ANR) REIDSOCs project (ANR 2016, Project ID: ANR-16-CE20-0010). ANR evaluated the scientific relevance and socio-economic impact of the study as part of the REIDSOCs project. ANR had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Commercial rams did not belong to any experimental design but were sampled by veterinarians and/or under veterinarian supervision for routine veterinary care and DNA collection. For the experimental animals (INRA, Domaine de La Fage), breeding conditions were similar to commercial sheep flocks. Blood collection and measurements followed procedures approved by the Regional Ethics Committee on Animal Experimentation, Languedoc-Roussillon (France), under the Agreement 752056/00.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France,

²Institut de l'Élevage, 31321 Castanet-Tolosan, France.

Received: 3 May 2019 Accepted: 29 August 2019

Published online: 18 September 2019

References

- Goddard ME, Kemper KE, MacLeod M, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc R Soc B Biol Sci.* 2016;283(1835):20160569. Cited 2019 Apr 18. <https://doi.org/10.1098/rspb.2016.0569>.
- Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics.* 1975;31(2):423–47. Cited 2019 Jun 26. Available from: <https://www.jstor.org/stable/2529430>.
- Fernando RL, Grossman M. Marker assisted selection using best linear unbiased prediction. *Genet Sel Evol.* 1989;21(4):467. Cited 2019 Jul 2. <https://doi.org/10.1186/1297-9686-21-4-467>.
- Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819–29. Cited 2018 Apr 13. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461589/>.
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 2009;92(9):4656–63. Cited 2018 Apr 30. Available from: <http://www.sciencedirect.com/science/article/pii/S0022030209707933>.
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 2010;42:2. Cited 2018 Apr 30. <https://doi.org/10.1186/1297-9686-42-2>.
- Chen CY, Misztal I, Aguilar I, Tsunuta S, Meuwissen THE, Aggrey SE, Wang T, Muir WM. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. *J Anim Sci.* 2011;109(1):23–8.
- Carillier C, Larroque H, Robert-Granié C. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. *Genet Sel Evol.* 2014;46(1):67. Cited 2019 Mar 28. <https://doi.org/10.1186/s12711-014-0067-3>.
- Yoshida GM, Carvalho R, Rodríguez FH, Lhorente JP, Yáñez JM. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics.* 2019;111(2):127–32. Cited 2019 Mar 28. Available from: <http://www.sciencedirect.com/science/article/pii/S0888754318300211>.
- Duchemin SI, Colonnari C, Legarra A, Baloché G, Larroque H, Astruc J-M, Barillet F, Robert-Granié C, Manfredi E. Genomic selection in the French Lacaune dairy sheep breed. *J Dairy Sci.* 2012;95(5):2723–33. Cited 2018 Apr 13. Available from: [http://www.journalofdairyscience.org/article/S0022-0302\(12\)00241-X/fulltext](http://www.journalofdairyscience.org/article/S0022-0302(12)00241-X/fulltext).
- Baloché G, Legarra A, Sallé G, Larroque H, Astruc J-M, Robert-Granié C, Barillet F. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J Dairy Sci.* 2014;97(2):1107–16.
- Astruc J-M, Baloché G, Buisson D, Labatut J, Lagriffoul G, Larroque H, Robert-Granié C, Legarra A, Barillet F. La sélection génomique des ovins laitiers en France. *INRA Prod Anim.* 2016;29(1):41–56. Cited 2018 Apr 13. Available from: <http://agris.fao.org/agris-search/search.do?recordID=FR2016228530>.
- Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet Res (Camb).* 2012;94(2):73–83.
- Zhang X, Lourenco D, Aguilar I, Legarra A, Misztal I. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of EBV and GWAS. *Front Genet.* 2016;7. Cited 2018 Apr 3. <https://doi.org/10.3389/fgenet.2016.00151/full>.
- Legarra A, Vitezica ZG. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. *Genet Sel Evol.* 2015;47:89. Cited 2018 Apr 3. <https://doi.org/10.1186/s12711-015-0165-x>.
- Teissier M, Larroque H, Robert-Granié C. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genet Sel Evol.* 2018;50(1). Cited 2018 Jun 25. <https://doi.org/10.1186/s12711-018-0400-3>.
- Teissier M, Larroque H, Robert-Granié C. Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *J Dairy Sci.* 2019;102(4):3142–54.
- Carillier-Jacquin C, Larroque H, Robert-Granié C. Including a s1 casein gene information in genomic evaluations of French dairy goats. *Genet Sel Evol.* 2016;48:54. Cited 2018 May 9. <https://doi.org/10.1186/s12711-016-0233-x>.
- Rupp R, Senin P, Sarry J, Allain C, Tascia C, Ligat L, Portes D, Woloszyn F, Bouchez O, Tabouret G, Lebastard M, Caubet C, Foucras G, Tosser-Klopp G. A point mutation in suppressor of cytokine signalling 2 (SOCS2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLOS Genet.* 2015;11(12):e1005629. Cited 2018 Apr 3. <https://doi.org/10.1371/journal.pgen.1005629>.
- Barillet F, Rupp R, Mignon-Grasteau S, Astruc J-M, Jacquin M. Genetic analysis for mastitis resistance and milk somatic cell score in French Lacaune dairy sheep. *Genet Sel Evol.* 2001;33(4):397–415. Cited 2018 Apr 3. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2705413/>.
- Rupp R, Lagriffoul G, Astruc JM, Barillet F. Genetic parameters for milk somatic cell scores and relationships with production traits in French Lacaune dairy sheep. *J Dairy Sci.* 2003;86(4):1476–81. Cited 2018 Apr 3. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0022030203737321>.
- Barillet F. Genetic improvement for dairy production in sheep and goats. *Small Rumin Res.* 2007;70(1):60–75. Cited 2018 Apr 11. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0921448807000090>.
- Marie-Etancelin C, Astruc JM, Porte D, Larroque H, Robert-Granié C. Multiple-trait genetic parameters and genetic evaluation of udder-type traits in Lacaune dairy ewes. *Livest Sci.* 2005;97(2):211–8. Cited 2019 Jan 29. Available from: <http://www.sciencedirect.com/science/article/pii/S0301622605001417>.
- Martin P, Pahiére I, Maroteau C, Clément V, David I, Klopp GT, Rupp R. Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. *J Dairy Sci.* 2018;101(6):5214–26. Cited 2019 Mar 21. Available from: [https://www.journalofdairyscience.org/article/S0022-0302\(18\)30261-3/abstract](https://www.journalofdairyscience.org/article/S0022-0302(18)30261-3/abstract).
- Oget C, Clément V, Pahiére I, Tosser-Klopp G, Fabre S, Rupp R. Genome-wide study finds a QTL with pleiotropic effects on semen and production traits in Saanen goats | request PDF. *Dubrovnik*; 2018. <https://prodinra.inra.fr/locale-fr/#ConsultNotice:450567>.
- Barillet F, Arranz J-J, Carta A. Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. *Genet Sel Evol.* 2005;37(Suppl 1):S109. Cited 2019 Apr 9. Available from: <http://www.gsejournal.org/content/37/S1/S109>.
- Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica.* 2009;136(2):245–57. Cited 2019 Mar 15. <https://doi.org/10.1007/s10709-008-9308-0>.
- Rupp R, Boichard D, Barbat A, Astruc JM, Lagriffoul G, Barillet F. Selection for mastitis resistance in French dairy sheep. In: *Proceedings of the 7th world congress on genetics applied to livestock production*. Montpellier; 2002. <http://www.wcgalp.org/system/files/proceedings/2002/selection-mastitis-resistance-french-dairy-sheep.pdf>.
- Barillet F, Astruc JM, Lagriffoul G, Aguerre X, Bonaiti B. Selecting milk composition and mastitis resistance by using a part lactation sampling design in French Manech red faced dairy sheep breed. In: *ICAR Technical Series*; 2009. p. 129–35. Cited 2018 Apr 3. Available from: <https://www.cabdirect.org/cabdirect/abstract/20103193070>.
- Kennedy BW, Quinton M, van Arendonk JA. Estimation of effects of single genes on quantitative traits. *J Anim Sci.* 1992;70(7):2000–12. Cited 2019 Apr 15. Available from: <https://academic.oup.com/jas/article/70/7/2000/4632001>.

31. Mota RR, Mayeres P, Bastin C, Glorieux G, Bertozzi C, Vanderick S, Hammami H, Colinet FG, Gengler N. Genetic evaluation for birth and conformation traits in dual-purpose Belgian blue cattle using a mixed inheritance model. *J Anim Sci*. 2017;95(10):4288–99. Cited 2019 Apr 29. Available from: <https://academic.oup.com/jas/article/95/10/4288/4771956>.
32. Martin P, Raoul J, Bodin L. Effects of the FecL major gene in the Lacaune meat sheep population. *Genet Sel Evol*. 2014;46:48. Cited 2018 May 9. <https://doi.org/10.1186/1297-9686-46-48>.
33. Sánchez A, Ilahi H, Manfredi E, Serradilla JM. Potential benefit from using the α S1-casein genotype information in a selection scheme for dairy goats. *J Anim Breed Genet*. 2005;122(1):21–9. Cited 2019 Apr 15. <https://doi.org/10.1111/j.1439-0388.2004.00474.x>.
34. Barillet F, Lagriffoul G, Marnet P-G, Laroque H, Rupp R, Portes D, Bocquier F, Astruc J-M. Objectifs de sélection et stratégie raisonnée de mise en oeuvre à l'échelle des populations de brebis laitières françaises. *INRA Prod Anim*. 2016;29(1):19–40. Cited 2018 Apr 13. Available from: <https://hal.archives-ouvertes.fr/hal-01519345>.
35. Barillet F. Amélioration génétique de la composition du lait des brebis : l'exemple de la race Lacaune. Paris: INA Paris-Grignon; 1985.
36. Ali AKA, Shook GE. An optimum transformation for somatic cell concentration in milk. *J Dairy Sci*. 1980;63(3):487–90. Cited 2018 May 17. Available from: <http://www.sciencedirect.com/science/article/pii/S0022030280829596>.
37. Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One*. 2009;4(3). Cited 2018 Apr 26. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2652362/>.
38. Brauning R, Carta A, Ciappesoni CG, Clarke S, Cockett N, Couldrey C, Daetwyler HD, Heaton MP, Kijas J, Larkin D, McCulloch A, McEwan J, McWilliam S, Moreno CR, Rowe S, Saunders G, Ventura R. Building the LD Chip and an update on the sheep genomes database. San Diego; 2016. Cited 2018 Dec 20. Available from: <https://pag.confex.com/pag/xxiv/webprogram/Paper22111.html>.
39. Larroque H, Chassier M, Saintilan R, Astruc J-M. Imputation accuracy from a low density SNP panel in 5 dairy sheep breeds in France. In: Oral presentation presented at: 68th annual meeting of the European Association of Animal Production. Tallinn; 2017. Available from: [http://www.eaap.org/Annual_Meeting/2017_tallinn/S\(09\)_01_Larroque.pdf](http://www.eaap.org/Annual_Meeting/2017_tallinn/S(09)_01_Larroque.pdf).
40. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15(1):478. Cited 2018 Sep 12. <https://doi.org/10.1186/1471-2164-15-478>.
41. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263–5.
42. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968;38(6):226–31.
43. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. Blupf90 and related programs (BGF90). In: Proceedings of the 7th world congress on genetics applied to livestock production. Montpellier; 2002. <http://www.wcgalp.org/system/files/proceedings/2002/blupf90-and-related-programs-bgf90.pdf>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



IV. Conclusions du Chapitre II.

Par cette étude, nous avons pu mettre en lumière que la fréquence de l'allèle mutant diminuait dans la population de mâles améliorateurs diffusés, potentiellement grâce à la sélection effectuée sur les cellules du lait dans l'objectif de sélection depuis 2005.

De plus, les méthodes pondérées testées ont apporté des gains de précisions sur les prédictions de l'ensemble des caractères en sélection, laissant envisager la possibilité d'une sélection plus efficace permettant un gain génétique plus rapide et conséquent sur la population globale. L'ajout du SNP *SOCS2* parmi les marqueurs de la puce a également un impact positif sur ces précisions, principalement sur le caractère SCS, bien que plus restreint en raison du déséquilibre de liaison de la mutation avec les marqueurs de la puce aux alentours.

Enfin, la méthode du *Gene Content* nous a semblé très intéressante dans la question de la gestion de mutations causales. En effet, cette méthode est un bon outil permettant de dissocier la valeur génétique d'un caractère en deux parts : une part due à l'effet du gène majeur, et une part due à l'effet de la partie polygénique restante. Cet outil peut donc être utilisé par les filières selon leur objectif de sélection. En l'occurrence, en brebis Lacaune, la sélection telle qu'elle est pratiquée aujourd'hui permet bien de diminuer la fréquence de l'allèle défavorable *SOCS2* tout en améliorant également la résistance aux mammites expliquée par le reste du génome.

En conclusion, cette étude nous a donc permis de confirmer que le gène *SOCS2* était bien pléiotrope, c'est-à-dire associé à différents caractères d'intérêt : fort effet sur l'inflammation mammaire, effets plus faibles mais néanmoins non négligeables sur les caractères de production laitière et de morphologie mammaire de profondeur de la mamelle. Ce dernier effet sur la morphologie mammaire, nouvellement étudié en ce qui concerne ce gène majeur, reflète probablement un effet indirect du gabarit et donc de la taille de l'animal. Les caractères étudiés dans ce chapitre sont les caractères en sélection, c'est-à-dire ceux qui sont disponibles à grande échelle et de grand intérêt pour la filière. D'autres caractères non encore en sélection sont tout de même importants tels que la croissance, ou encore la capacité de reproduction. Nous allons dans le chapitre suivant étudier l'effet de la mutation *SOCS2* sur de nouveaux caractères non encore en sélection en routine.

Chapitre III. Association du génotype *SOCS2* aux mammites et aux caractères de production dans des populations ovines expérimentales et commerciales

I. Objectifs et démarches

Nous venons de voir que la mutation du gène *SOCS2* à l'étude influence plusieurs caractères en sélection tels que la production laitière, l'inflammation et la morphologie mammaires, qui sont des phénotypes disponibles à grande échelle grâce au dispositif d'enregistrement du schéma de sélection. Dans notre premier chapitre bibliographique, nous avons vu que la protéine *SOCS-2* était impliquée dans d'autres phénomènes biologiques tels que la réponse immunitaire, la reproduction, ou encore la croissance.

Nous avons donc cherché à évaluer dans un premier temps dans ce chapitre les effets de cette mutation en ovin sur la réponse infectieuse, c'est-à-dire le nombre de bactéries dans le lait, qui est un critère important pour la filière en terme de qualité sanitaire du lait. De nouveaux phénotypes relatifs à la santé mammaire (bactériologie du lait, examens cliniques) étaient disponibles dans une unité expérimentale INRA (La Fage) pour des brebis génotypées avec une puce à SNPs spécifiquement construite sur des régions QTL associées aux mammites qui contenait le SNP de la mutation *SOCS2*. Nous avons réalisé des études d'associations pangénomiques (GWAS, *Genome-Wide Association Studies*) sur ce jeu de données afin de mettre en lumière une association de *SOCS2* avec la réponse infectieuse, mais aussi afin de valider une nouvelle fois l'association de *SOCS2* avec les caractères de production. Un jeu de données d'une autre race (Manech Tête Rousse), également génotypée avec cette puce, a pu aussi être analysée (caractères de production laitière et inflammation mammaire). Ces résultats seront présentés dans ce chapitre dans les deux parties suivantes sous forme d'un article publié (communication courte). Cette communication courte a été acceptée pour publication dans la revue *Genetics Selection Evolution* le 1^{er} février 2019.

Ensuite, il nous paraissait également important d'évaluer l'effet de cette mutation sur la reproduction qui représente un coût important pour la filière et qui n'est pas encore pris en compte en sélection. Des phénotypes femelles (taux de réussite à l'IA et prolificité) étaient disponibles pour des brebis de la ferme expérimentale La Fage génotypées pour la mutation

SOCS2. Des modèles mixtes linéaires généralisés ont été réalisés pour quantifier précisément cet effet.

Enfin, l'effet de *SOCS2* sur la croissance et le gabarit n'avait pas encore pu être observé directement précédemment dans cette thèse, et, étant donné l'importance de la protéine SOCS-2 dans la régulation de l'hormone de croissance (gigantisme chez la souris), ces phénotypes étaient également intéressants à évaluer. Nous souhaitons notamment évaluer l'effet de cette surcroissance sur la morphologie (présence éventuelle de tares). Des phénotypes étaient disponibles dans les stations de testage où les jeunes béliers sont pesés et pointés sur différents postes de morphologie à leur entrée dans les centres. En effet, les mâles possédant des tares morphologiques sont éliminés, ce qui représente un surcoût pour le schéma. Des modèles mixtes linéaires ont été réalisés pour quantifier l'effet de *SOCS2* sur ces différents phénotypes.

Les résultats non publiés (reproduction et croissance) seront présentés dans ce chapitre (partie IV) sous le format classique des articles scientifiques.

II. Effet de la mutation *SOCS2* sur les caractères de production et le statut infectieux (Oget *et al.*, 2019a)

II.1. Introduction

L'identification de *loci* ou de mutations causales associés à la résistance aux mammites peut être utile pour les schémas de sélection en brebis laitière comme ça l'est déjà en bovin à travers le monde. Sept régions génomiques associées au caractère CCS, un indicateur indirect de l'infection de la mamelle, ont déjà été identifiées en ovin (race espagnole Churra, race française Lacaune, et une population issue du croisement entre les races Sarde et Lacaune, respectivement italienne et française). Dans cette étude, nous avons utilisé une puce à SNP ovine comprenant 960 marqueurs, sur les races de brebis laitières Lacaune et Manech Tête Rousse afin de valider ces sept régions génomiques associées aux mammites.

II.2. Résultats et discussion

Le SNP le plus significatif (rs868996547) sur le chromosome OAR 3 est la mutation précédemment décrite dans le gène *SOCS2*. Un effet antagoniste entre santé et croissance de ce gène candidat causal chez la brebis Lacaune a été confirmé. Les effets de la mutation sur l'état infectieux de la mamelle, c'est-à-dire l'augmentation du nombre de cellules somatiques et l'excrétion de bactéries dans le lait, ont également été identifiés. Ce SNP n'était pas présent dans les génotypages disponibles en race Manech Tête Rousse. Trois autres régions associées aux mammites ont également été confirmées sur les chromosomes OAR16 (Manech Tête Rousse), 19 (Lacaune) et 2 (les deux races). Pour la région OAR2, nous avons validé les SNP précédemment détectés dans plusieurs autres races (Sarda, Churra et Chios). Pour les SNP significatifs dans ces quatre régions QTL associées aux mammites, l'effet variait de 0,24 à 0,67 exprimé en écart-type phénotypique des caractères. Deux des régions QTL associées aux mammites (OAR2 et 16) validées ici sont également associées de manière antagoniste aux caractères de production laitière dans les deux populations.

II.3. Conclusion

Ces résultats révèlent, du moins en partie, une base génomique pour le compromis entre production laitière et résistance aux mammites. Quatre des sept régions QTL associées aux mammites qui étaient auparavant identifiées dans des populations indépendantes, ont été confirmées dans cette étude, ce qui démontre un partage partiel des mécanismes génétiques liés aux mammites entre différentes populations éloignées de brebis laitières.

III. Communication courte (Oget *et al.*, 2019a)

SHORT COMMUNICATION

Open Access



A validation study of loci associated with mastitis resistance in two French dairy sheep breeds

Claire Oget^{1*}, Charlotte Allain¹, David Portes², Gilles Foucras³, Alessandra Stella⁴, Jean-Michel Astruc⁵, Julien Sarry¹, Gwenola Tosser-Klopp^{1†} and Rachel Rupp^{1†}

Abstract

Background: The identification of loci associated with resistance to mastitis or of the causative mutations may be helpful in breeding programs for dairy sheep as it is for cattle worldwide. Seven genomic regions that control milk somatic cell counts, an indirect indicator of udder infection, have already been identified in sheep (Spanish Churra, French Lacaune and Italian Sardinian–Lacaune backcross populations). In this study, we used a 960 custom-designed ovine single nucleotide polymorphism (SNP) chip in Lacaune and Manech Tête Rousse dairy sheep to validate these seven genomic regions associated with mastitis.

Results: The most significant SNP (*rs868996547*) on *Ovis aries* chromosome (OAR) 3 was a previously described mutation in the *suppressor of cytokine signalling 2* (*SOCS2*) gene. An antagonist effect of this causal candidate between health and growth in Lacaune sheep was confirmed. Effects of the mutation on the infectious status of the udder, i.e. increases in milk somatic cell counts and bacteria shedding, were also identified. This SNP was not present in the data available on Manech Tête Rousse. Three other regions associated with mastitis were also confirmed on OAR16 (Manech Tête Rousse), 19 (Lacaune) and 2 (both breeds). For the OAR2 region, we validated previously detected SNPs in several other breeds (Sarda, Churra, and Chios). For significant SNPs in the four mastitis regions, the effect varied from 0.24 to 0.67 phenotypic standard deviation of the traits. Two of the mastitis quantitative trait loci (QTL) regions (OAR2 and 16) that we validated here were also associated in opposite ways with milk production traits in both populations.

Conclusions: These results indicate, at least in part, a genomic basis for the trade-off between milk production and mastitis resistance. Four of the seven mastitis QTL regions that were previously identified in independent populations, were confirmed in this study, which demonstrates partial sharing of mastitis-related genetic mechanisms between different distant dairy sheep populations.

Background

Mastitis is an inflammation of the mammary gland, which in dairy sheep is mostly due to bacterial infections by *Staphylococci* [1]. Mastitis is a serious burden for the milk industry due to the altered quality of milk and increased cost of flock renewal. Beside hygienic measures, genetic selection for improved resistance to mastitis

is now implemented in breeding programs for several breeds of dairy ruminants worldwide [2]. However, its application to dairy sheep is still rare, mainly because the recording cost per animal, relative to potential income, is prohibitive for many traits other than production traits. In sheep, the identification of loci that are associated with resistance to udder infection or the causative mutations may be helpful in selection. However, resistance to mastitis is highly complex and the genetically determined biological basis behind this trait remains unknown.

Several quantitative trait loci (QTL) regions that control milk somatic cell count (SCC), an indirect indicator

*Correspondence: claire.oget@inra.fr

[†]Gwenola Tosser-Klopp and Rachel Rupp contributed equally to this work

¹GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France
Full list of author information is available at the end of the article



© The Author(s) 2019. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

of udder infection, have been identified in dairy sheep through the EU-funded 3SR project (Sustainable solutions for small ruminants, FP7-KBBE-245140) [3]. For one of these QTL, Rupp et al. [4] identified a single nucleotide polymorphism (SNP) in the coding frame of the *suppressor of cytokine signalling 2 (SOCS2)* gene as the putative causal mutation associated with high SCC in the Lacaune breed. A few QTL regions were then confirmed by Banos et al. [5] in a population of the Greek Chios breed using four mastitis indicator traits, namely clinical mastitis occurrence, milk SCC, total viable bacterial count in milk and the California mastitis test.

The objective of our study was to confirm the ovine QTL that control mastitis resistance in two independent dairy sheep populations, using a 960 custom-designed ovine SNP chip.

Methods

Two independent French dairy sheep populations were used: Lacaune ewes (N=504) from a divergent selection based on extreme breeding values for SCC at the experimental facility of La Fage (INRA, UE 321, Roquefort, France) [6], and Manech Tête Rousse rams (N=145) raised in the CDEO (Ordiarp, France) testing station in 2013 (birth year from 2008 to 2011). Among the 504 Lacaune individuals, 213 ewes belonged to the high SCC line (42.2%) and 291 to the low SCC line (57.7%). The selection lines were about three genetic standard deviations (SD) apart [6].

In the Lacaune ewe population, milk yield, fat content, protein content, and SCC were measured monthly at morning milking. Test-day SCC were log-transformed for normality into SCS [7]. The arithmetic averages of the first lactation test-days were then computed and corrected for year of sampling for fat content (FAT_L1), protein content (PROTEIN_L1), and SCS (LSCS_L1). The milk yield (MILK_L1) trait was computed as the trait used for genetic evaluation, i.e. the 250-day cumulative production adjusted for lactation length and standardized to an adult production ($\times 1.3$). MILK_L1 was multiplied by 1.3 to follow the definition used in the Manech Tête Rousse breed for genetic evaluation and allow direct comparison of average milk production between both breeds.

Staphylococcus spp. abundance in milk was measured at three-time points during the first lactation by a qPCR-based technique developed at the "Interactions Hôtes – Agents Pathogènes" (IHAP) laboratory (Toulouse, France). Briefly, milk was collected aseptically from each half udder independently after precleaning and disinfecting the teat apex using a cotton wool moistened with 70% alcohol. Whole milk was centrifuged (6000g, 20 min) before two consecutive enzymatic proteolytic treatments

with lysozyme and proteinase K. DNA was extracted using a Biosprint 96 semi-robotic workstation and DNA Blood kits (QIAGEN), and finally eluted in 50 μ L of distilled water. An internal DNA control (QIAGEN) was used to assess recovery and lack of qPCR inhibitors. High-throughput qPCR in 384-well format was performed on 1 μ L of DNA extract in a total volume of 5 μ L using *tuf*-specific primers (*tuf* 5'-CAC GAC CAG TGA TTG AGA ATA CG and *tuf* 3'-CCA ATG CCA CAA ACT CGT GA), probe (CCA TTC ATG ATG CCA GTT G), and the Quantifast Pathogen PCR kit (QIAGEN). The proportion of inhibited samples was lower than 5%. Values above the cycle threshold were compared to a standard curve obtained from known amounts of genomic DNA from a *Staphylococcus aureus* laboratory strain and expressed as a bacterial titre (quantity of equivalent bacterial genomes per volume of milk), on a logarithmic scale. The three results were averaged for each ewe and corrected for the effects of month and year of sampling (STAPH_L1).

Chronic mastitis was based on the presence of mammary abscesses, recorded by clinical examination (ABSCESS_L1). Animals were noted as "1" (case) when the presence of at least one abscess was detected at least twice, whereas animals were noted as "0" (control) when they were found to be healthy (without any abscess) at least three times during the first lactation.

Each ewe was weighed at birth (W_BIRTH), at 100 days (W_DAY_100) and 250 days (W_DAY_250), after the first (on average 412 days, W_1ST_LAMB) and second lambing (on average 744 days, W_2ND_LAMB), and at the age of 920 days (W_DAY_920). Phenotypes were corrected for year and feeding method (breastfeeding or artificial suckling). Basic statistics are in Table 1.

In the Manech Tête Rousse population, SCC and milk production traits were obtained from the official milk records. In this breed, milk yield is measured monthly and SCC, fat and protein contents are measured three to four times during the first three lactations [8]. For association mapping, we used the daughter yield deviations (DYD) [9] from regular national genetic evaluations for milk production traits (MILK, PROTEIN, and FAT) and lactation average somatic cell scores (LSCS). DYD correspond to the average performance of the daughters of a ram, corrected for the environmental effects and the genetic value of the dams (Table 2).

Both Lacaune and Manech Tête Rousse populations were genotyped with a 960 custom-designed ovine SNP chip [10]. The chip was designed and developed within the 3SR EU project [3] based on several QTL for SCC that were previously identified in Spanish Churra [11], French Lacaune [4] and Italian Sardinian–Lacaune backcross populations ([12] and personal

Table 1 Number of animals with a phenotype (N) and basic statistics for the Lacaune population of 504 ewes

Trait	N	Mean (\pm SD)
LSCS_L1	494	3.1 \pm 1.6
FAT_L1 (g/L)	494	64.1 \pm 7.1
PROTEIN_L1 (g/L)	494	53.5 \pm 3.4
MILK_L1 ^a (L)	493	334.9 \pm 61.9
STAPH_L1	404	0.7 \pm 0.8
ABSCESS_L1	376	15 cases/361 controls
W_BIRTH (kg)	502	3.9 \pm 0.6
W_DAY_100 (kg)	492	28.7 \pm 3.0
W_DAY_250 (kg)	472	50.6 \pm 4.1
W_1ST_LAMB (kg)	438	64.7 \pm 6.7
W_2ND_LAMB (kg)	393	73.4 \pm 7.7
W_DAY_920 (kg)	378	67.3 \pm 7.2

^a MILK_L1 is the 250-day cumulative production adjusted for lactation length and standardized to an adult production (\times 1.3)

Table 2 Number of animals with a phenotype (N) and basic statistics for the Manech Tête Rousse population of 145 rams and their 15,722 offspring

Trait	Rams ^a		Offspring ewes ^b	
	N	Mean (\pm SD)	N	Mean (\pm SD)
LSCS	145	0.1 \pm 0.6	15,115	3.6 \pm 1.5
FAT (g/L)	145	0.9 \pm 4.2	15,368	61.9 \pm 8.4
PROTEIN (g/L)	145	0.2 \pm 2.6	15,368	49.3 \pm 4.4
MILK ^c (L)	145	84.7 \pm 25.2	15,722	270 \pm 79

^a Phenotypes are DYD expressed as deviations from the average population used for genetic evaluation

^b Phenotypes are the observed performances during the first lactation

^c MILK is the 250-day cumulative production adjusted for lactation length and standardized to an adult production (\times 1.3)

communication). Using these previous association studies, seven regions of interest (Table 3) on *Ovis aries* (OAR) chromosomes 2, 3, 5, 16 and 18 were selected based on commonalities among populations found at the time (in 2012) or on their high significance. SNPs were selected within these regions from the 54 K or 800 K Illumina ovine chips [13] or from novel genome sequencing within the 3SR project. The 10 SNPs in the OAR3 region included the causal mutation in the *SOCS2* gene and nine other closely linked loci that had been identified by Rupp et al. [4]. Genomic positions refer to the ovine reference genome v3.1 [14]. After quality control, the following SNPs were excluded from the analyses: non-polymorphic SNPs, SNPs with a missingness rate higher than 5%, with a minor allele frequency lower than 2%, and SNPs that deviated from

Hardy–Weinberg proportion ($p < 1E-05$), thus, 745 and 708 SNPs were selected for the Lacaune and Manech Tête Rousse populations, respectively.

Genome-wide association studies (GWAS) were performed for each phenotype using the polygenic univariate mixed model approach implemented in the genome-wide efficient mixed-model association (GEMMA) software [15]. The polygenic effect was fitted using a covariance structure according to the genomic relationship matrix. Corrections were applied to account for multiple testing. First, a Bonferroni correction of $\alpha = 5\%$ was applied (significance threshold = α /number of SNPs). SNPs with a p value less than $6.6E-05$ and less than $7.1E-05$ were considered as highly significantly associated for the Lacaune and the Manech Tête Rousse populations, respectively. Since association tests are not independent, due to the large number of SNPs in high linkage disequilibrium within the QTL regions, and to several traits being highly correlated, a less restrictive suggestive significance threshold was also calculated (significance threshold = $[\alpha/\text{number of independent regions in the chip}]/\text{number of independent variables in the study}$). Table 3 lists the seven independent regions that were found to contain such SNPs on the 960 custom-designed ovine SNP chip. We used two methods to obtain the number of independent variables: a factor analysis of mixed data (FAMD) for the Lacaune population, for which we had to consider quantitative (LSCS_L1, FAT_L1, PROTEIN_L1, MILK_L1, STAPH_L1, weights) and qualitative (ABSCESS_L1) phenotypes, and a principal component analysis (PCA) for the Manech Tête Rousse population, for which all phenotypes were quantitative. FAMD and PCA using phenotypes were computed using the FactoMineR package [16] of the R software [17]. According to the clustering elbow method, the number of independent variables is the marginal point where the percentage of variance explained by the PCA dimensions drops and produces an angle in the histogram. This method led us to choose the first five dimensions of the FAMD (Lacaune) and the first two dimensions of the PCA (Manech Tête Rousse), which explained 77.5% (Lacaune) and 79.6% (Manech Tête Rousse) of the variance. Therefore, we used $N1 = 5$ (Lacaune) and $N2 = 2$ (Manech Tête Rousse), the number of independent variables in the study, leading to suggestive significance thresholds of $1.4E-03$ and $3.6E-03$, respectively.

Results and discussion

Significant SNPs from the GWAS are in Table 4 (Lacaune) and Table 5 (Manech Tête Rousse). The first noteworthy result concerns the highly significant region on OAR3 in the Lacaune population. Indeed, three SNPs, which were associated with mastitis and growth traits,

Table 3 SNPs of the 3SR-mastitis-960-SNP custom chip, selected within the 3SR EU project for their association with mastitis-related traits

OAR	Discovery population	QTL ^a (Mb)	Interval (Mb)		Origin of the SNP			Number of SNPs		
			Start	End	54 K chip	800 K chip	SEQ	Before QC	After QC LAC	After QC MTR
2	LAC	125	124.6	125.5	19	10		29	24	24
2	SAR	206.7	204.2	209.7	106	146		252	196	180
	CHU	208.7								
3	LAC	129.7	129.7	130.1	4	6	10	20	13	12
5	LAC	96.5	95.8	96.2	11	9		20	17	16
12	CHU	18.3	17.9	20.4	54	76		130	101	97
	SAR	18.4								
16	LAC	36.2	28.3	36.8	144	110		254	202	183
19	CHU	26.3	23.6	28.9	75	180		255	202	196
	LAC	28.6								

OAR *Ovis aries* chromosome, LAC Lacaune, SAR Sardinian–Lacaune backcross, CHU Churra, QTL quantitative trait locus, Mb megabase, SEQ sequence, QC quality control, MTR Manech Tête Rousse

^a QTL location

Table 4 List of SNPs associated with the different phenotypes for the Lacaune population genotyped with the 3SR-mastitis-960-SNP custom chip

OAR	SNP ID	SNP position	MAF %	Estimation of SNP effect	SE	p value of Wald test	Significant	Trait
2	rs414463145	204256340	23	−2.1	0.6	1.1E−03		FAT_L1
	rs420100221	204333224	32	−2.0	0.6	6.9E−04		FAT_L1
	rs409223720	206118959	46	0.0	0.0	1.3E−03		ABSCCESS_L1
	rs406554330	206311785	35	0.1	0.0	8.2E−04		ABSCCESS_L1
	rs421656393	207088920	34	2.0	0.6	1.2E−03		FAT_L1
	rs409595486	209135680	12	2.6	0.8	7.6E−04		FAT_L1
3	<i>rs428428896</i>	<i>129680367</i>	22	<i>0.6</i>	<i>0.1</i>	<i>9.6E−06</i>	***	LSCS_L1
			22	0.3	0.1	1.4E−03		STAPH_L1
			21	2.0	0.6	1.1E−03		W_1ST_LAMB
	rs425616833	129685397	39	−0.4	0.1	3.2E−04		LSCS_L1
	<i>rs868996547</i>	<i>129722200</i>	17	<i>0.8</i>	<i>0.1</i>	<i>3.0E−07</i>	***	LSCS_L1
			17	0.3	0.1	6.8E−04		STAPH_L1
			17	1.0	0.3	6.0E−04		W_DAY_100
			17	1.5	0.4	3.7E−04		W_DAY_250
			16	2.4	0.7	5.0E−04		W_1ST_LAMB
			15	2.8	0.9	9.5E−04		W_DAY_920
	<i>rs426941860</i>	<i>129927538</i>	20	<i>0.7</i>	<i>0.1</i>	<i>5.2E−07</i>	***	LSCS_L1
			20	0.8	0.3	1.2E−03		W_DAY_100
			19	2.4	0.7	1.0E−03		W_DAY_920
16	rs404369966	28321220	25	−0.2	0.0	3.2E−04		W_BIRTH
	<i>rs403769730</i>	<i>35801129</i>	6	<i>34.5</i>	<i>8.4</i>	<i>5.2E−05</i>	***	MILK_L1
			6	1.5	0.4	3.4E−04		W_DAY_100
			7	2.1	0.6	6.8E−04		W_DAY_250
19	rs412825949	25639818	32	−0.3	0.1	4.1E−04		STAPH_L1
	<i>rs416350585</i>	<i>25684229</i>	14	<i>3.6</i>	<i>0.9</i>	<i>3.4E−05</i>	***	W_DAY_920
	rs400011699	28086791	38	−14.9	4.0	2.6E−04		MILK_L1

SNPs in italic are significant with the Bonferroni threshold and the others are significant with the suggestive threshold

OAR *Ovis aries* chromosome, MAF minor allele frequency, SE standard error of the estimated SNP effect

Table 5 List of SNPs associated with the different phenotypes for the Manech Tête Rouse population genotyped with the 3SR-mastitis-960-SNP custom chip

OAR	SNP ID	SNP position	MAF %	Estimation of SNP effect	SE	p value of Wald test	Trait
2	rs403180214	206231638	50	-0.3	0.1	1.7E-04	LSCS
	rs430641657	207218842	12	16.0	4.3	2.9E-04	MILK
	rs425558509	208650955	42	-1.5	0.5	2.6E-03	FAT
	rs398918249	209409342	43	-1.0	0.3	1.3E-03	PROTEIN
16	rs404369966	28321220	26	1.0	0.3	3.6E-03	PROTEIN
			26	-10.2	3.4	2.7E-03	MILK
	rs428277629	28526130	35	1.0	0.3	2.0E-03	PROTEIN
	rs418639132	28632456	41	1.0	0.3	1.0E-03	PROTEIN
	rs405638287	29030021	37	1.1	0.3	7.4E-04	PROTEIN
	rs399031726	29072599	36	1.2	0.3	2.5E-04	PROTEIN
	rs422927329	29304504	37	1.2	0.3	1.1E-04	PROTEIN
	rs402269497	30319608	19	-12.5	3.8	1.2E-03	MILK
	rs421638047	30692759	28	1.9	0.6	1.1E-03	FAT
			28	1.2	0.3	6.1E-04	PROTEIN
			28	-12.9	3.4	2.1E-04	MILK
	rs409981325	31298560	10	0.4	0.1	1.9E-03	LSCS
	rs409954614	34424863	37	1.0	0.3	1.2E-03	PROTEIN

There are no significant SNPs at the Bonferroni threshold for this breed, all the SNPs presented in this table are significant with the suggestive threshold
OAR *Ovis aries* chromosome number, MAF minor allele frequency, SE standard error of the estimated SNP effect

were detected at the Bonferroni threshold. The most significant SNP (*rs868996547*, p value = $3.0E-07$) was the causal mutation in the *SOCS2* gene, previously reported by Rupp et al. [4]. This mutation causes a loss in functional activity of the *SOCS2* protein, which is involved in inflammatory response control and growth [18] through the JAK/STAT/SOCS pathway. The lowest p values and highest estimates of effects for this SNP were observed for both mastitis traits (LSCS_L1 and STAPH_L1) and four of the six weight traits. Corresponding effects varied from 0.33 SD for W_DAY_100 to 0.50 SD for LSCS_L1. Thus, we confirmed an adverse effect of the *SOCS2* gene point mutation on mammary inflammation and growth, as reported by Rupp et al. [4]. We also found that the mutation had an unfavourable effect on the infectious status of the udder (0.38 SD), since the low-frequency allele increased cell counts and bacteria shedding in milk. All these results confirm the pleiotropic effect of the *SOCS2* mutation on body growth and the host's control of mastitis. This SNP did not segregate in the Manech Tête Rouse population although 12 other SNPs segregated in this narrow genomic region (Table 3). No QTL for mastitis was detected in this region for the Manech Tête Rouse population (Table 5), which provides further evidence that *rs868996547* is a strong candidate in Lacaune but is absent from Manech Tête Rouse. Moreover, we found that there was no effect of the SNPs of

the same region on OAR3 when the SNP considered as causal (*rs868996547*) was included as a fixed factor in the model for the Lacaune population analyses. Indeed, we observed an increase of the p values of the SNPs that surround the mutation for all traits for which the association was previously significant. For example, the minimum p value for the LSCS_L1 trait in the OAR3 region was $1.5E-02$ (*rs425616833*), which confirmed that the other SNPs in the region do not explain any additional variance.

Then, we applied suggestive thresholds, which allowed us to confirm three other regions that are associated with mastitis. In the Lacaune population, regions on OAR2 and 19 are associated with the ABSCESS_L1 and STAPH_L1 traits, respectively, and regions on OAR2 and 16 are also significant for the LSCS trait in the Manech Tête Rouse population. For significant SNPs, the effect varied from 0.38 SD (OAR19 in Lacaune) to 0.67 SD (OAR16 in Manech Tête Rouse) of the traits (Tables 4, 5). These three QTL regions had already been identified in Sarda (OAR2), Churra (OAR19 and 2) (Table 3) and Chios breeds (OAR2, 16 and 19) [5]. Thus, for these regions, and especially OAR2, our data reinforce the hypothesis of true mastitis QTL, which might involve similar genes and pathways across breeds. Banos et al. [5] suggested several candidate genes for OAR2: *cytotoxic T-lymphocyte-associated protein 4* (*CTLA4*);

204,777,523–204,784,522 bp), *inducible T-cell co-stimulator* (*ICOS*; 204,851,429–204,873,693 bp), and *isocitrate dehydrogenase 1 (NADP+)*, *soluble* (*IDH1*; 209,236,699–209,259,307 bp), which are involved in T-cell responses or mitigation of oxidative damage. However, the region between 204.2 and 209.7 Mbp on OAR2 includes more than 90 genes, which makes the identification of appropriate candidates difficult.

Interestingly, two of the mastitis QTL regions (OAR2 and 16) that were confirmed in the present study, were also associated with milk production traits in both populations. Moreover, the region on OAR16 is strongly (Bonferroni threshold) associated with MILK_L1 (effect = 0.56 SD) in the Lacaune population and with MILK, FAT, and PROTEIN (effect = 0.50 SD, 0.45 SD and 0.46 SD, respectively) in the Manech Tête Rousse population. Thus, the underlying QTL could be a QTL for milk production that has an indirect impact on mastitis. In Lacaune, the positive sign of the estimated effects of SNP *rs403769730* shows that this QTL is favourable for milk production (Table 4), i.e. leading to an increase in milk quantity, but unfavourable for LSCS_L1 (results not shown), i.e. leading to an increase in somatic cell count. A similar pattern is observed for SNP *rs421638047* in Manech Tête Rousse, for which the estimated effects have a negative sign, i.e. leading to a decrease in MILK which is unfavourable for milk yield (Table 5) and a decrease in LSCS which is favourable for the health of the animal (results not shown). These results are in agreement with the positive and antagonistic correlation that exists between mastitis and milk production trait in Lacaune [19] and other ovine [20] and bovine breeds [21], which indicates, at least in part, a genomic basis for the trade-off between milk production and mastitis resistance.

Conclusions

We confirmed four out of seven QTL regions for mastitis in the Lacaune population, and only two in the Manech Tête Rousse population. This is consistent with the fact that Lacaune belongs to the breeds for which these regions were first discovered, although the individuals were different. The two significant regions detected in the Manech Tête Rousse population are rather encouraging, unlike a similar study on nematode resistance where QTL validation was inconclusive [22]. These results demonstrate that mastitis-related genetic mechanisms are shared between different distant dairy sheep populations.

Authors' contributions

CO performed the association analyses, contributed to their interpretation and wrote the draft. CA and DP conducted the experiment at the facility of La Fage, collected and prepared the data from the Lacaune population. GF developed and carried out the measurements of PCR-based milk bacteriology in the Lacaune population. He also led the REIDSOCs ANR funded project. AS, GTK and RR developed the 35R-mastitis-960-SNP chip. JMA contributed

to data collection and performed phenotype calculations in the Manech Tête Rousse population. JS conducted the DNA extractions. GTK and RR designed the study and helped to interpret the analyses. All authors read and approved the final manuscript.

Author details

¹ GenPhySE, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France. ² INRA, Domaine de La Fage, 12250 Saint-Jean-et-Saint-Paul, France. ³ IHAP, Université de Toulouse, INRA, ENVT, Castanet-Tolosan, France. ⁴ Parco Tecnologico Padano (PTP), Via Einstein, Cascina Codazza, 26900 Lodi, Italy. ⁵ Institut de l'Élevage, 31321 Castanet-Tolosan, France.

Acknowledgements

The authors are grateful to Francis Fidelle (CDEO, Ordiarp, France) and the technical staff of La Fage (INRA, France) for assisting with animal care and data collection. We thank Edinburgh Genomics (formerly ARK Genomics platform) for genotyping the 960-SNP chip. Thanks are also extended to Julie Demars (INRA Toulouse, France) for her help in handling blood and DNA samples, and Christian Tasca for preparing milk samples for bacteriology by qPCR. The work was funded by the 7PCRD EU 35R project (Sustainable Solutions for Small Ruminants, FP7-KBBE-245140). Claire Oget acknowledges the support of the Agence Nationale de la Recherche (ANR) for her scholarship (Project Reidsocs, ID: ANR-16-CE20-0010).

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Information about the 960 custom-designed ovine SNP chip that was developed within WP3 of the 35R program and used for the analyses in the current study is available online: http://genoweb.toulouse.inra.fr/~tosser/35R-WP3-960_snp_mastitis/. The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The commercial rams from the Manech Tête Rousse population did not belong to any experimental design but were sampled by veterinarians and/or under veterinarian supervision for routine veterinary care and DNA collection. For the experimental animals (INRA, Domaine de La Fage), breeding conditions were similar to those of commercial sheep flocks. Blood collection and measurements followed procedures approved by the Regional Ethics Committee on Animal Experimentation, Occitanie (France), under the Agreement 752056/00.

Funding

This work was supported by grants from the EU 35R project (FP7-KBBE, Project ID: 245140) and the French National Research Agency (ANR) REIDSOCs project (ANR-2016, Project ID: ANR-16-CE20-0010).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 September 2018 Accepted: 1 February 2019

Published online: 13 February 2019

References

- Bergonier D, de Crémoux R, Rupp R, Lagriffoul G, Berthelot X. Mastitis of dairy small ruminants. *Vet Res*. 2003;34:689–716.
- Rupp R, Foucras G. Genetics of mastitis in dairy ruminants. In: Bishop SC, Axford RFE, Nicholas FW, Owen JB, editors. *Breeding for disease resistance in farm animals*. 3rd ed. Wallingford: CAB; 2010. p. 183–212.
- Sustainable solutions for small ruminants. Projects FP7-KBBE European Commission. CORDIS European Commission. https://cordis.europa.eu/project/rcn/95054_en.html. Accessed 17 May 2018.

4. Rupp R, Senin P, Sarry J, Allain C, Tasca C, Ligat L, et al. A point mutation in *suppressor of cytokine signalling 2* (*Socs2*) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet*. 2015;11:e1005629.
5. Banos G, Bramis G, Bush SJ, Clark EL, McCulloch MEB, Smith J, et al. The genomic architecture of mastitis resistance in dairy sheep. *BMC Genomics*. 2017;18:624.
6. Rupp R, Bergonier D, Dion S, Hygonenq MC, Aurel MR, Robert-Granié C, et al. Response to somatic cell count-based selection for mastitis resistance in a divergent selection experiment in sheep. *J Dairy Sci*. 2009;92:1203–19.
7. Ali AKA, Shook GE. An optimum transformation for somatic cell concentration in milk. *J Dairy Sci*. 1980;63:487–90.
8. Barillet F, Astruc JM, Lagriffoul G, Aguerre X, Bonaiti B. Selecting milk composition and mastitis resistance by using a part lactation sampling design in French Manech red faced dairy sheep breed. In: Proceedings of the 36th ICAR biennial session: 16–20 June 2008. Niagara Falls; 2009. p. 129–35. <https://www.cabdirect.org/cabdirect/abstract/20103193070>. Accessed 3 Apr 2018.
9. VanRaden PM, Wiggins GR. Derivation, calculation, and use of national animal model information. *J Dairy Sci*. 1991;74:2737–46.
10. Index of http://genoweb.toulouse.inra.fr/~tosser/35R-WP3-960_snp_mastitis/. Accessed 17 May 2018.
11. Gutiérrez-Gil B, Esteban-Blanco C, Suarez-Vega A, Arranz JJ. Detection of quantitative trait loci and putative causal variants affecting somatic cell score in dairy sheep by using a 50 K SNP-chip and whole-genome sequencing. *J Dairy Sci*. 2018;101:9072–88.
12. Sechl S, Casu S, Casula M, Congiu GB, Mlari S, Mulas G, et al. Genome-wide association analysis of resistance to paratuberculosis and mastitis in dairy sheep. In: Proceedings of the 64th annual meeting of the European federation of animal science: 26–30 August 2013. Nantes; 2013. http://old.eaap.org/Previous_Annual_Meetings/2013Nantes/Papers/Published/554_06.pdf. Accessed 17 May 2018.
13. Nicolazzi EL, Caprera A, Nazzicari N, Cozzi P, Strozzi F, Lawley C, et al. SNPchip v3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*. 2015;16:283.
14. Ovis aries – Ensembl genome browser 92. http://www.ensembl.org/Ovis_aries/info/index. Accessed 17 May 2018.
15. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012;44:821–4.
16. Lê S, Josse J, Hussen F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25:1–18.
17. R Development Core Team. R: A language and environment for statistical computing. 2005. <http://www.R-project.org/>. Accessed 1 Jun 2018.
18. Metcalf D, Greenhalgh CJ, Viney E, Willson TA, Starr R, Nicola NA, et al. Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature*. 2000;405:1069–73.
19. Rupp R, Lagriffoul G, Astruc JM, Barillet F. Genetic parameters for milk somatic cell scores and relationships with production traits in French Lacaune dairy sheep. *J Dairy Sci*. 2003;86:1476–81.
20. Tolone M, Riggio V, Portolano B. Estimation of genetic and phenotypic parameters for bacteriological status of the udder, somatic cell score, and milk yield in dairy sheep using a threshold animal model. *Livest Sci*. 2013;151:134–9.
21. Rupp R, Boichard D. Genetics of resistance to mastitis in dairy cattle. *Vet Res*. 2003;34:671–88.
22. Keane OM, Hanrahan JP, McRae KM, Good B. An independent validation study of loci associated with nematode resistance in sheep. *Anim Genet*. 2018;49:265–8.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



IV. Analyses complémentaires

Dans un second temps, nous avons évalué l'effet de *SOCS2* sur la reproduction et sur la croissance des animaux. Nous allons donc voir dans cette partie des analyses d'association sur de nouveaux caractères femelles (reproduction) et mâles (morphologie, croissance).

IV.1. Effet de la mutation *SOCS2* sur la reproduction femelle

IV.1.a. Matériels et méthodes

IV.1.a.i. Animaux et phénotypes

Nous avons travaillé sur des données de reproduction de femelles provenant de l'unité expérimentale INRA La Fage. Ces données correspondent à 4 815 saillies de 1 398 brebis de millésimes de naissance entre 2011 et 2018 sur 8 campagnes de reproduction (2012 à 2019). Dans cette unité expérimentale, les brebis sont mises à la reproduction au mois d'août dans un premier temps par Insémination Artificielle (IA), sauf cas exceptionnels lors de protocoles expérimentaux où par exemple 140 saillies en première intention ont été réalisées en monte naturelle. Lorsque l'IA n'a pas fonctionné (vérifié grâce à des échographies réalisées par les techniciens de l'unité), les brebis sont mises une seconde fois à la lutte, en monte naturelle cette fois-ci. Les nombres de saillies disponibles dans ce jeu de données par type (IA ou monte naturelle) pour la première et seconde intentions sont présentés dans le Tableau 3.

Tableau 3 Description des saillies en première et seconde intentions des femelles de La Fage sur les campagnes de 2012 à 2019

<i>Type de saillie</i>	<i>Nombre de saillies</i>	
	<i>1^{ère} saillie</i>	<i>2^{nde} saillie</i>
<i>IA</i>	3 599	0
<i>MN</i>	140	1 076

Abréviations : IA : Insémination artificielle ; MN : Monte Naturelle.

Ces brebis sont toutes génotypées pour la mutation *SOCS2*. Sont disponibles 942 homozygotes sauvages C/C, 409 hétérozygotes C/T, et 47 homozygotes mutants T/T. Le premier phénotype étudié est la réussite à la saillie, et plus particulièrement à l'IA qui représente 96% des saillies en première intention (3 599 phénotypes disponibles pour 1 373 brebis). Ce caractère est binaire (0 ou 1), où 1 signifie que la saillie est fécondante, et 0, que la saillie est non fécondante. Le second phénotype étudié est le nombre de petits par portée quand la première IA est fécondante

(2 560 phénotypes pour 1 168 brebis). La description des effectifs des saillies selon les génotypes *SOCS2* est présentée dans le Tableau 3.

Tableau 4 Répartition des effectifs de saillies selon les génotypes *SOCS2*

<i>Génotype SOCS2</i>	<i>Réussite à la 1^{ère} IA (1 373 brebis)</i>		<i>Taux de réussite à la 1^{ère} IA</i>	<i>Répartition du nombre de petits par portée à la 1^{ère} IA fécondante (1 168 brebis)</i>				
	Oui	Non		1	2	3	4	5
	<i>C/C</i>	1 810		679	73%	22%	61%	14%
<i>C/T</i>	707	331	68%	22%	59%	16%	2%	1%
<i>T/T</i>	43	29	60%	16%	77%	7%	0%	0%
Total	2 560	1 039	71%	22%	61%	15%	2%	0%

Abréviations : IA : Insémination artificielle ; MN : Monte Naturelle.

IV.1.a.ii. Analyses de variance

Toutes les analyses statistiques présentées ici ont été réalisées avec le logiciel R version 3.5.0 (R Development Core Team, 2005). Des analyses de variance ont été effectuées pour déterminer les effets du génotype *SOCS2* sur les différents phénotypes présentés précédemment.

Les deux phénotypes étudiés (réussite à l'IA et nombre de petits par portée) sont des données répétées car les saillies sont disponibles pour plusieurs campagnes de reproduction d'un même animal. Nous avons donc appliqué des modèles mixtes pour étudier l'effet du génotype *SOCS2* sur ces deux caractères. Le phénotype de réussite à l'IA étant un caractère binaire, et le nombre de petits par portée pouvant être assimilé à une variable de comptage, nous avons choisi d'appliquer des modèles linéaires généralisés à effets mixtes (fonction `glmer()` du package R *lme4* version 1.1.17 (Bates *et al.*, 2015)), avec respectivement pour le premier et le second phénotype des distributions modèles binomiale logistique et de Poisson. En plus de l'effet aléatoire animal, ces modèles incluaient un effet fixe combiné du numéro de lactation de la brebis mise à la saillie et de l'année de la campagne de reproduction. Les brebis de lactations supérieures ou égales à 3 (disponibles jusqu'à 7 lactations maximum initialement) ont été regroupées.

IV.1.b. Résultats et discussion

Sur l'ensemble des campagnes de 2012 à 2019, on peut tout d'abord constater (Tableau 4) que le taux de réussite à l'IA est plus élevé chez les individus C/C (73%) que chez les individus C/T (68%) et que chez les individus T/T (60%). La mutation aurait donc un effet négatif sur la réussite à l'IA. La régression logistique réalisée sur le caractère binaire de réussite à l'IA (1 = réussite ; 0 = échec) a confirmé cela avec un effet significatif du génotype *SOCS2* (F-valeur globale = 4,89, correspondant à des p-valeurs de 0,026 et 0,012 respectivement pour les effets des génotypes C/T et T/T par rapport au génotype de référence C/C). Cet effet du génotype sur la réussite à l'IA est également significatif lorsqu'on regarde les risques relatifs d'apparition du phénotype. En effet, ce risque est de 0,94 et de 0,77, respectivement pour les C/T et T/T (intervalles de confiance respectifs [0,88 ; 0,99] et [0,583 ; 0,953]), par rapport aux C/C (référence). Cela signifie que les brebis possédant au moins un allèle mutant ont moins de probabilité de réussite à l'IA (la valeur 1 n'est pas comprise dans les intervalles de confiance), et donc, autrement dit, plus de risque d'échec (1,3 fois plus de risque d'échec à l'IA pour les T/T par exemple avec un intervalle de confiance de [1,05 ; 1,71]).

Les résultats obtenus pour le second phénotype ont montré que la mutation *SOCS2* n'a pas d'effet significatif sur le nombre de petits par portée (F-valeur globale pour l'effet génotype de 0,08). Ceci pouvait être observé grâce à la répartition du nombre de petits par génotype *SOCS2* (Tableau 4). En effet, en comparant les C/C et les C/T (les T/T sont trop peu nombreuses pour être comparées ici), on peut voir une répartition similaire : une grande proportion de portée (61% chez les C/C et 59% chez les C/T) sont doubles, un peu moins d'un quart des femelles mettent bas un seul petit (22% des portées pour les deux génotypes), puis 14% (C/C) à 16% (C/T) ont trois petits, et enfin moins de 5% ont plus de 3 petits par portée.

La protéine SOCS-2 interviendrait donc dans le phénomène de reproduction, sans effet sur le nombre de petits lorsqu'il y a eu fécondation. Nous avons vu dans le [Chapitre I](#) qu'en bovin, les protéines SOCS intervenaient au moment de l'implantation de l'embryon grâce notamment à la reconnaissance maternelle acquise lors de ce processus (Carvalho *et al.*, 2014). Le jeu de données disponible utilisé comportait quelques indications concernant l'échec de la saillie (avortement, vide, portée morte, etc.) mais n'était pas assez complet pour réaliser une analyse sur cet aspect. Des analyses complémentaires sur ce sujet seraient intéressantes à poursuivre notamment dans le cadre de l'unité expérimentale La Fage où un suivi plus fin des brebis mise à la reproduction pourrait être réalisé. De nombreuses études s'intéressent au parallélisme entre phénomène d'ovulation et phénomène inflammatoire qui impliquerait une régulation par la voie JAK/STAT. Par exemple, chez l'homme, il a été montré que dans les cellules de la granulosa,

des gènes liés à l'inflammation étaient sur-exprimés au moment de l'ovulation, initiée par des voies de signalisation telles que la voie NF-KB (*nuclear factor-kappa B*), la voie MAPK (*Mitogen-activated protein kinases*), ou encore la voie JAK/STAT (Poulsen *et al.*, 2019). En bovin, l'implication directe de la protéine régulatrice SOCS-2 a été suggérée par sa surexpression avant l'ovulation en comparaison à après l'ovulation et à la phase lutéale, coïncidant avec une augmentation de la concentration de l'IGF (Mense *et al.*, 2015). Le taux d'ovulation chez les brebis portant les différents génotypes *SOCS2* pourrait être ainsi intéressant à observer.

IV.1.c. Conclusion

En conclusion, l'effet de la mutation *SOCS2* sur la capacité de reproduction des individus femelles est confirmé par ces analyses avec un effet significatif sur le taux de réussite à l'IA. Ce résultat est cohérent avec la bibliographie qui a montré qu'en bovin et en ovin, les protéines SOCS participent à la régulation du phénomène d'implantation embryonnaire. La reproduction étant un mécanisme complexe, des analyses complémentaires sont nécessaires pour déterminer précisément à quel stade biologique (ovulation, fécondation, implantation et développement embryonnaires, etc.) la protéine SOCS-2 pourrait intervenir.

IV.2. Effet de la mutation *SOCS2* sur la croissance et la morphologie fonctionnelle du mâle

IV.2.a. Matériels et méthodes

IV.2.a.i. Animaux et phénotypes

Les phénotypes proviennent de 7 260 béliers entrés dans les deux centres d'élevage en Lacaune lait (4 391 mâles d'Ovitest, et 2 869 de Confédération), au cours des trois campagnes 2015 à 2017. Les béliers entrent en centre d'élevage à environ 1 mois d'âge (novembre-décembre) et sont pesés trois fois par les techniciens agréés des centres : une fois à leur entrée, une fois en janvier, et une fois en mars. Des valeurs de Poids à Age Type (PAT) à 35, 60, 90, 120 et 150 jours, ainsi que des Gains Moyens Quotidiens (GMQ) entre 35 et 60, entre 60 et 90, entre 90 et 120, et entre 120 et 150 jours, sont ensuite déduits par interpolation de ces mesures.

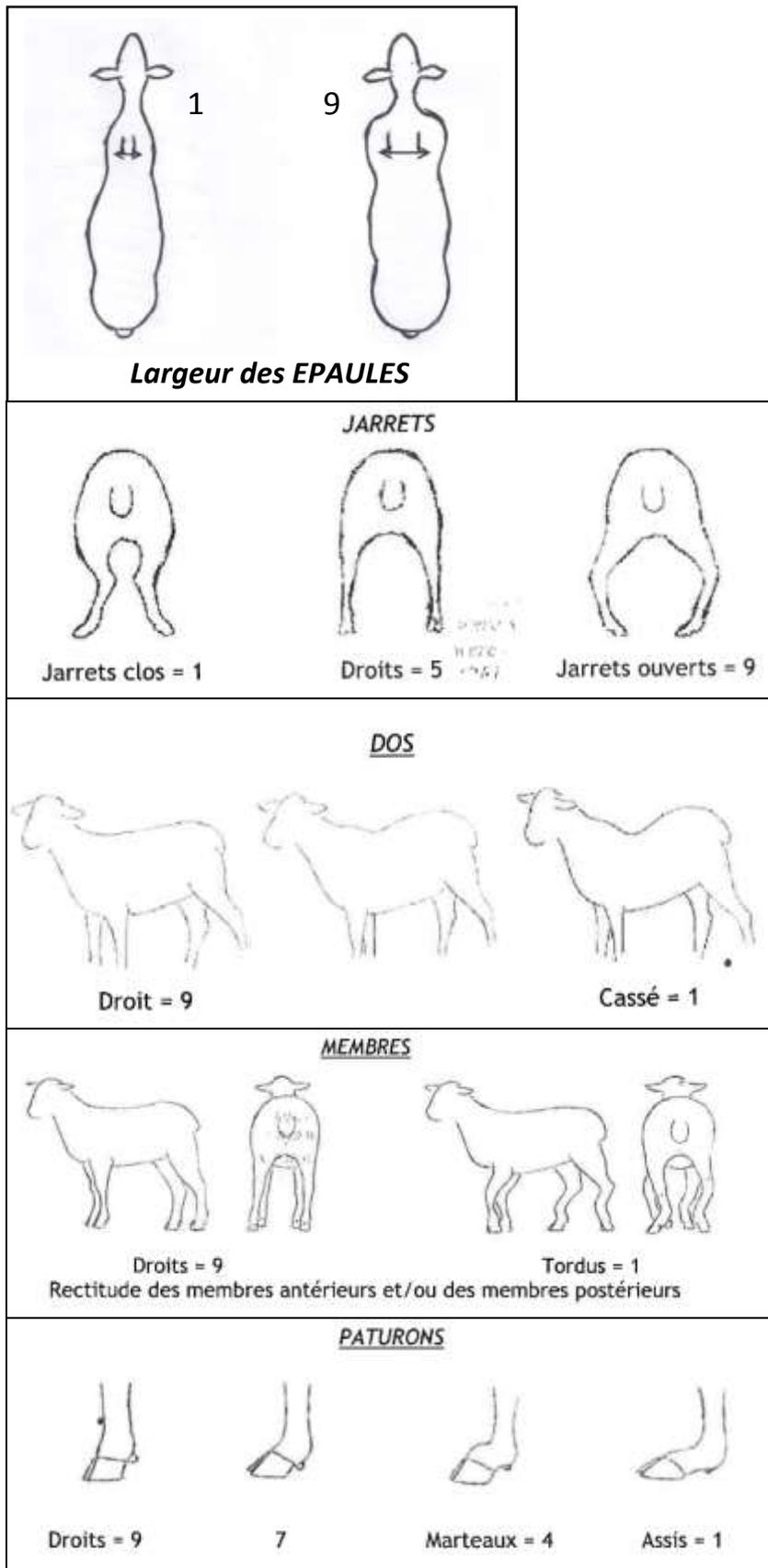


Figure 10 Description schématique des sept phénotypes de morphologie fonctionnelle du mâle, évalués dans les centres d'élevage (Jean-Michel Astruc, communication personnelle)

Ces animaux sont également pointés sur leur morphologie fonctionnelle au moment de la troisième pesée (en moyenne 5 mois d'âge) sur sept postes (Figure 10) : largeur des épaules (phénotype « epaule »), orientation des jarrets (« jarret »), ligne de dos (« ligdos »), rectitude des membres postérieurs (« membar ») et antérieurs (« membav »), et forme/longueur des pâturons arrière (« patuar ») et avant (« patuav »). Ces postes sont évalués selon une échelle linéaire de scores de 1 à 9 donnés par les techniciens qualifiés de chaque centre. Les entreprises de sélection réalisent ces pointages afin de ne pas entrer en centre d'IA les béliers futurs améliorateurs possédant des défauts morphologiques. Actuellement, environ 40% des béliers sont éliminés pour défaut morphologique (Jean-Michel Astruc, communication personnelle), ce qui représente un coût pour les entreprises de sélection. Une évaluation génétique (polygénique) de ces postes est réalisée chaque année et diffusée aux entreprises de sélection (cf. programme CASDAR Maxi'mâle, rédaction finale en cours). Une évaluation génomique est prévue afin de réaliser un pré-choix encore plus précoce des béliers (avant leur phénotypage). Au vu de l'effet de *SOCS2* sur la croissance (syndrome du gigantisme chez la souris), il est probable qu'une surcroissance des individus entraîne des malformations physiques, d'où l'intérêt d'évaluer l'effet de *SOCS2* sur les données de morphologie fonctionnelle du mâle disponibles.

Nous avons joint ces phénotypes aux génotypes *SOCS2* qui étaient disponibles pour la majorité de ces animaux (vrais ou imputés d'après le chapitre précédent). Le facteur génotype *SOCS2* possède trois niveaux : les homozygotes sauvages C/C, les hétérozygotes C/T, et les homozygotes mutants T/T. Les effectifs, moyennes et écart-types des différents phénotypes sont présentés dans le Tableau 5.

Tableau 5 Description des phénotypes (morphologie, croissance) des mâles génotypés *SOCS2* : effectifs et statistiques descriptives

Phénotype	Effectif avec génotype <i>SOCS2</i> connu				Moyenne ± écart-type	
	Total	T/T	C/T	C/C		
<i>PAT35 (kg)</i>					15,60 ± 2,38	
<i>PAT60 (kg)</i>					22,59 ± 2,69	
<i>PAT90 (kg)</i>					30,81 ± 3,47	
<i>PAT120 (kg)</i>					38,53 ± 3,87	
<i>PAT150 (kg)</i>	2 782	17	602	2 163	46,21 ± 4,58	
<i>GMQ3560 (g/j)</i>					279,71 ± 61,78	
<i>GMQ6090 (g/j)</i>					273,82 ± 45,50	
<i>GMQ90120 (g/j)</i>					257,31 ± 43,11	
<i>GMQ120150 (g/j)</i>					255,96 ± 43,88	
<i>epaule</i>					6,84 ± 1,13	
<i>jarret</i>					4,38 ± 0,88	
<i>ligdos</i>					7,20 ± 1,12	
<i>membar</i>	6 455	26	1 304	5 125	7,44 ± 0,91	
<i>membav</i>					7,20 ± 1,09	
<i>patuar</i>					7,34 ± 1,18	
<i>patuav</i>					6,87 ± 1,31	

Abréviations : *GMQ* : Gain Moyen Quotidien ; *PAT* : Poids à Âge Type ; *epaule* : largeur des épaules ; *jarret* : orientation des jarrets ; *ligdos* : ligne de dos ; *membar* : rectitude des membres postérieurs ; *membav* : rectitude des membres antérieurs ; *patuar* : forme/longueur des pâturons arrière ; *patuav* : forme/longueur des pâturons avant.

IV.2.a.ii. Analyses de variance

Toutes les analyses statistiques présentées ici ont été réalisées avec le logiciel R version 3.5.0 (R Development Core Team, 2005). Des analyses de variance ont été effectuées pour déterminer les effets du génotype *SOCS2* sur les différents phénotypes présentés précédemment.

Les mesures de croissance (PAT et GMQ) étant des données répétées qui suivent des lois normales, nous avons appliqué des modèles linéaires mixtes (fonction `lme()` du package R *nlme* version 3.1.137 (Pinheiro *et al.*, 2019)). Ces modèles incluaient un effet aléatoire animal (données répétées 5 fois pour les PAT et 4 fois pour les GMQ), ainsi que les effets fixes suivants :

- Âge
 - ✓ 35, 60, 90, 120 et 150 jours pour les PAT
 - ✓ Entre 30 et 60 jours, entre 60 et 90 jours, entre 90 et 120 jours, et entre 120 et 150 jours pour les GMQ
- Génotype *SOCS2* (C/C, C/T et T/T)
- Interaction Âge/génotype
- Campagne : 2015 et 2016
- Centre d'élevage : Confédération et Ovitest.

Une structure de corrélation sur les mesures répétées par individus a été appliquée entre les différents temps à l'aide d'un processus autorégressif d'ordre 1. Les effets des génotypes intra-temps (comparaisons entre les niveaux de génotypes deux à deux intra-temps, c'est-à-dire tests *post-hoc*) ont ensuite été effectuées, avec une correction pour les tests multiples à l'aide de la procédure de Bonferroni (fonction `lsmeans()` du package R *lsmeans* v2.23 (Lenth, 2016)).

Les pointages étant des mesures non répétées, nous avons appliqué des modèles linéaires afin d'évaluer l'effet du facteur génotype *SOCS2* (fonction `lm()`). Les données étant légèrement dissymétriques, nous avons effectué une transformation logarithmique sur les données brutes ($\log_{10}(10-x)$) pour s'approcher au mieux de la distribution normale (à l'exception du phénotype jarret qui avait déjà une distribution proche de la normalité). Les effets fixes du modèle sont les suivants : génotypes *SOCS2*, campagne, centre d'élevage, âge au phénotypage et pointeur. Les comparaisons entre les niveaux de génotypes deux à deux (tests *post-hoc*) ont ensuite été effectuées de la même manière que précédemment.

IV.2.b. Résultats et discussion

Les résultats des modèles mixtes linéaires sur les phénotypes de croissance sont présentés dans le Tableau 6. L'effet du génotype *SOCS2* sur les PAT était significatif, confirmant ainsi l'effet de la mutation sur le poids des individus. Cet effet se traduit par une augmentation du poids chez les individus C/T en comparaison des individus C/C, et de manière de plus en plus significative à partir de 90 jours (p-valeur C/C - C/T à 90j = 0.0188, à 120j = 0.0013, et à 150j

= 0.0001). Les individus C/T ont en effet un poids augmenté par rapport aux C/C en moyenne (lsmeans) de 1,43%, 1,47% et 1,48% respectivement pour PAT90, PAT120 et PAT150. Les différences observées avec les individus T/T ne sont pas significatives en raison du trop faible effectif de ces individus dans ce jeu de données (Tableau 5). Cet effet sur le poids avait déjà été observé auparavant sur 18 brebis (Rupp *et al.*, 2015). Cette étude montrait que la différence de poids entre C/C et T/T était croissante au cours du temps et significative à partir de 530 jours d'âge, et que la différence entre individus homozygotes à 3 ans d'âge était de 18%. Notre hypothèse pour expliquer la différence de pourcentage de gain de poids entre cette étude et la nôtre serait que les individus étudiés ici sont très jeunes et que la différence s'accentuerait avec l'âge. En revanche, nous n'avons pas observé d'effet du génotype *SOCS2* sur les GMQ. Ceci peut s'expliquer par le fait que le GMQ est une mesure qui reste à une échelle de temps relativement faible (moyenne d'un gain par jour) tandis que l'effet de la mutation s'exprime à plus long terme. Il serait intéressant d'étudier un gain sur une période plus longue afin de confirmer notre hypothèse de croissance plus rapide en lien avec le poids de plus en plus élevé chez les individus porteurs de la mutation.

Les résultats des modèles linéaires sur les phénotypes de morphologie sont présentés dans le Tableau 7. Nous avons observé quatre phénotypes dont des associations significatives étaient mises en lumière avec la mutation *SOCS2* : la largeur des épaules, la ligne de dos, la rectitude des membres postérieurs et la forme/longueur des pâturons arrière. On distingue trois cas différents d'association : (1) l'effet *SOCS2* global est significatif mais les comparaisons deux à deux des génotypes ne montrent aucune différence significative (rectitude des membres postérieurs et forme/longueur des pâturons arrière), (2) l'effet global *SOCS2* n'est pas significatif mais on observe une différence significative entre les génotypes en les comparant deux à deux (ligne de dos), et (3) l'effet global *SOCS2* est significatif, ainsi que les comparaisons deux à deux (largeur des épaules).

Concernant le premier cas d'association, on observe une tendance de scores plus faibles chez les mâles porteurs de la mutation, signifiant des membres et pâturons respectivement plus tordus et assis. Pour ce résultat, nous n'avons pas d'explication claire, ceci est observé pour la première fois. Cependant, une hypothèse est envisageable : la différence de moyennes entre le groupe des individus C/C et C/T et les individus T/T est grande (-0,28 et -0,37 respectivement pour « membar » et « patuar »), même en regard des erreurs d'estimation dues au très faible effectif des T/T (respectivement 0,17 et 0,22), tandis que les C/C et les C/T ne sont pas différents (respectivement 7,56 et 7,58 ± 0,02 pour « membar » et 7,54 ± 0,02 et 7,52 ± 0,03 pour « patuar »). La significativité du test global serait donc due aux T/T qui sont mal estimés en

raison d'un trop faible effectif. Ce résultat est donc à prendre avec précaution et nécessite validation avec un effectif plus important d'individus T/T.

Le deuxième cas d'association met en lumière une différence significative de score (« ligdos ») entre les individus C/C et C/T. En effet, les C/T possèdent un score en moyenne (lsmeans) plus élevé que les C/C ($7,35 \pm 0,05 < 7,44 \pm 0,05$) et une tendance également à la hausse en moyenne est observée chez les T/T ($7,54 \pm 0,19$). Cela signifie que les individus porteurs de la mutation auraient une ligne de dos moins rectiligne que les individus non porteurs. Ce résultat est également observé pour la première fois et serait à confirmer avec un jeu de données plus conséquent. En effet, si la mutation est à l'origine de malformations morphologiques, ces phénotypes pourraient être pris en considération dans les évaluations.

Le dernier cas d'association (« epaule ») est plus classique, avec une significativité globale et également au niveau des comparaisons deux à deux, et est plus facilement interprétable au vu des précédents résultats obtenus dans la littérature. En effet, on observe une différence significative entre la moyenne (lsmeans) des T/T ($6,00 \pm 0,29$) et celle des C/T ($6,83 \pm 0,07$) et également avec celle des C/C ($6,90 \pm 0,06$). Cela signifie que les individus porteurs de la mutation ont un score de largeur des épaules plus faible et seraient donc moins larges. Au premier abord, ce résultat semble incohérent avec la littérature (Rupp *et al.*, 2015) qui a montré que les individus porteurs de mutation sont plus grands et plus larges. Cependant, contrairement aux résultats des auteurs (Rupp *et al.*, 2015) qui avaient à disposition des mesures précises des distances entre les différents os, notre phénotype dans cette étude est une appréciation globale de la morphologie sous forme de score notée de 1 à 9. Notre première hypothèse pour expliquer ce résultat surprenant serait donc la suivante : les individus porteurs de la mutation étant plus grands et plus lourds (confirmé par nos analyses sur les PAT précédemment), et certainement plus longilignes (les individus sont jeunes donc probablement peu engraisés), on peut avoir tendance à les considérer plus fins dans l'appréciation globale du gabarit (sans notion de mesure), d'où un score de largeur des épaules plus faible pour les individus porteurs de la mutation. Une deuxième hypothèse apparaît également pour notre jeu de données : les différences significatives sont observées uniquement vis-à-vis des T/T pour lesquels nous avons observé précédemment des poids bien plus légers en moyenne que pour les autres individus mais avec une grande erreur d'estimation, probablement en raison du faible effectif de ces individus. Des analyses supplémentaires avec un effectif plus important et des mesures plus précises (mesures plutôt que notes) sont donc nécessaires pour confirmer ces résultats.

Tableau 6 Résultats des modèles linéaires mixtes effectués sur les mesures de croissance des mâles des stations

Phénotype	P-valeur (effet SOCS2)	Âge (j)	C/C (lsmean ± erreur standard)	C/T (lsmean ± erreur standard)	T/T (lsmean ± erreur standard)
PAT (kg)	0,001 **	35	15,19	15,50	14,89
		60	22,18	22,49	21,97
		90	30,37 ^b ± 0,08	30,80 ^a ± 0,14	30,26 ^{ab} ± 0,83
		120	38,06 ^b	38,62 ^a	37,79 ^{ab}
		150	45,71 ^b	46,39 ^a	45,29 ^{ab}
GMQ (g/j)	0,4181	35_60	279,20	279,38	282,14
		60_90	272,52	276,32	275,26
		90_120	256,02	260,09	249,79
		120_150	254,72	258,53	248,79

Abréviations : GMQ : Gain Moyen Quotidien ; PAT : Poids à Âge Type. Les astérisques en exposant (**) signifient que l'effet du génotype SOCS2 dans le modèle est significatif avec $p < 0,01$. Les lettres différentes en exposant (^{ab}) montrent la différence significative entre les génotypes avec $p < 0,05$, $p < 0,01$ et $p < 0,001$ respectivement pour PAT90, PAT120 et PAT150.

Tableau 7 Résultats des modèles linéaires effectués sur les données de pointages des mâles des stations

Phénotype	P-valeur (effet SOCS2)	C/C			C/T			T/T		
		(lsmean ± erreur standard)			(lsmean ± erreur standard)			(lsmean ± erreur standard)		
<i>epaule</i>	3.633E-07 ***	6.90	^a	± 0.06	6.83	^a	± 0.07	6.00	^b	± 0.29
<i>jarret</i>	0.351	4.35		± 0.04	4.35		± 0.05	4.48		± 0.17
<i>ligdos</i>	0.190	7.35	^a	± 0.05	7.44	^b	± 0.05	7.54	^{ab}	± 0.19
<i>membar</i>	0.019 *	7.56		± 0.02	7.58		± 0.02	7.28		± 0.17
<i>membav</i>	0.184	7.27		± 0.02	7.33		± 0.03	7.08		± 0.19
<i>patuar</i>	0.008 **	7.54		± 0.02	7.52		± 0.03	7.16		± 0.22
<i>patuav</i>	0.497	7.12		± 0.03	7.17		± 0.04	7.06		± 0.23

Abréviations : *epaule* : largeur des épaules ; *jarret* : orientation des jarrets ; *ligdos* : ligne de dos ; *membar* : rectitude des membres postérieurs ; *membav* : rectitude des membres antérieurs ; *patuar* : forme/longueur des pâturons arrière ; *patuav* : forme/longueur des pâturons avant. Les astérisques signifient que l'effet du génotype SOCS2 dans le modèle est significatif avec $p < 0,01$ (*), $p < 0,05$ (**), et $p < 0,001$ (***). Les lettres différentes en exposant (^{ab}) montrent la différence significative entre les génotypes.

IV.2.c. Conclusion

Pour conclure sur l'association de *SOCS2* aux caractères de croissance et de morphologie des mâles, nous avons pu voir que la mutation était associée aux poids (PAT) grâce à un grand jeu de données d'homozygotes sauvages C/C et d'hétérozygotes C/T, avec un poids supérieur chez les individus porteurs de la mutation. Concernant les caractères de morphologie, les résultats sont à nuancer car très probablement influencés par le trop faible effectif d'homozygotes mutés T/T dans le jeu de données. Il semblerait cependant qu'il y ait une association avec quelques phénotypes de rectitude des membres qu'il serait intéressant d'étudier plus finement dans le futur. Une autre limitation de cette étude est l'âge des individus disponibles (quelques mois d'âge). En effet, étant donné qu'il a été montré précédemment dans la littérature que l'effet de *SOCS2* s'accroît avec l'âge, nous soupçonnons une sous-évaluation de l'effet du gène *SOCS2* par rapport à son effet sur l'ensemble de la carrière de ces béliers. Des mesures supplémentaires sont donc nécessaires pour conclure sur cette partie.

V. Conclusions du Chapitre III.

Dans ce chapitre, nous cherchions à évaluer quantitativement les effets de *SOCS2* sur les caractères d'intérêt pour la filière et susceptibles d'être influencés par la mutation *SOCS2* : statut infectieux, reproduction, croissance, production. Pour ce faire, nous avons utilisé différents jeux de données (commerciaux, expérimentaux) et différentes approches statistiques (« tout-génome » et modèles linéaires avec uniquement le génotype *SOCS2* en effet fixe).

Nous avons commencé par réaliser des GWAS avec une puce sur laquelle le SNP *SOCS2* était disponible. Ces GWAS ont confirmé l'effet de *SOCS2* dans une population de brebis Lacaune sur les caractères de production laitière et d'inflammation et d'infection mammaires, avec notamment de nouveaux phénotypes de la réponse infectieuse (bactériologie, examens cliniques). De plus, la mutation n'a pas été trouvée dans la population de Manech Tête Rousse étudiée (mâles des stations de testage). Ces GWAS ont également mis en lumière d'autres régions potentiellement pléiotropes et partagées entre les différentes races de brebis laitières.

Les autres caractères disponibles et intéressants à étudier étaient des caractères croissance et de morphologie mâle. Des modèles linéaires ont montré pour les phénotypes mâles que la mutation était associée à une augmentation de poids. Cependant cette étude est limitée par l'effectif trop faible d'individus T/T et l'âge des individus (jeunes), raisons pour lesquelles il nous est difficile de conclure quant aux phénotypes de morphologie. Des données supplémentaires sont nécessaires.

Concernant les caractères de reproduction femelle, des modèles linéaires généralisés ont montré une forte association de la mutation avec le taux de réussite à l'IA et aucune association avec la prolificité des brebis. Des analyses complémentaires sont nécessaires pour déterminer précisément à quel stade biologique (ovulation, fécondation, implantation et développement embryonnaires, etc.) la protéine SOCS-2 pourrait intervenir.

Ce chapitre ouvre donc de nombreuses perspectives d'études quant à l'association de cette mutation à de nouveaux caractères, et notamment celui de la reproduction qui est de grand intérêt pour la filière. En effet, un projet d'évaluation sur la fertilité femelle est actuellement en cours d'élaboration en vue d'une insertion prochaine de ce caractère dans le schéma de sélection.

Chapitre IV. Réponse à une infection expérimentale chez des brebis portant les différents génotypes *SOCS2*

I. Objectifs et démarches

Nous venons de voir à l'échelle populationnelle quels étaient les effets de la mutation R96C ovine du gène *SOCS2* sur les différents caractères quantitatifs d'intérêt pour la filière. Les caractères les plus influencés par la mutation étaient liés aux mammites et par conséquent à la réponse infectieuse : présence de cellules somatiques et de bactéries dans le lait. Afin de déterminer plus précisément le rôle de la protéine SOCS-2 dans les mécanismes biologiques sous-jacents, j'ai étudié la réponse de brebis en fonction de la présence de la mutation R96C dans le cadre d'une infection contrôlée.

Pour ce faire, j'ai conduit une expérimentation basée sur une infection intra-mammaire par un staphylocoque (*S. aureus*) chez des brebis portant les génotypes *SOCS2* extrêmes, produites au sein de la ferme expérimentale INRA La Fage. Nous avons réalisé un suivi clinique fin de ces brebis ainsi que de nombreux prélèvements de lait et de sang au cours de l'infection pour la caractérisation moléculaire de la réponse immunitaire. De nombreuses données phénotypiques, cliniques et transcriptomiques ont ainsi été générées.

Ce chapitre présentera, sous le format d'un article scientifique, les résultats non encore publiés de l'analyse de cette expérience (modèles linéaires, analyses d'expression des gènes et analyses multivariées permettant l'intégration de l'ensemble des données).

II. Matériel et méthodes

Cette expérimentation animale, prévue dans le projet ANR REIDSOCS (<https://anr.fr/Projet-ANR-16-CE20-0010>), a fait l'objet d'une saisine (APAFIS#12333-2017112618176260) validée par le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation, afin de justifier de l'utilisation des animaux à des fins scientifiques.

Dans cette partie seront présentés les choix réalisés concernant :

- les individus échantillonnés pour représenter les différents génotypes de la mutation à l'étude dans le gène *SOCS2*,
- les échantillons prélevés,
- les modalités de collecte et de traitement des échantillons,
- les analyses de laboratoire effectuées sur les différents échantillons,
- les traitements bio-informatiques et les analyses statistiques réalisés sur les variables générées.

II.1. Plan expérimental

II.1.a. Choix des brebis

D'après les résultats obtenus précédemment dans cette population (Rupp *et al.*, 2015), les différences de phénotype observées étaient maximales entre les individus homozygotes C/C et T/T. Par ailleurs, l'effet de la mutation semble proche de l'additivité. Nous avons donc décidé de comparer ces deux génotypes C/C et T/T.

Pour des raisons pratiques (gestion des animaux au cours du protocole), éthiques (respect de la règle de réduction du nombre d'animaux utilisés à des fins scientifiques), scientifiques (nombre d'animaux suffisant pour l'analyse statistique), et économiques (coût de la gestion des animaux et des analyses à effectuer sur les échantillons prélevés), nous avons retenu 16 individus (8 C/C et 8 T/T) pour cette expérimentation (Tableau 8). Cet échantillon d'animaux, que l'on souhaite représentatif de la race Lacaune, provient de l'unité expérimentale INRA La Fage.

Tableau 8 Description des brebis initialement choisies (toutes les lignes) et retenues (lignes non grisées) pour les infections expérimentales

<i>Numéro brebis</i>	<i>Date de naissance brebis (année 2016)</i>	<i>Génotype SOCS2 brebis</i>	<i>Poids brebis (kg)</i>	<i>Nombre de fœtus (gestation brebis)</i>	<i>Génotype SOCS2 mère</i>	<i>Père</i>	<i>Génotype SOCS2 père</i>
472	05/12	C/C	46.8	1	C/C	1	C/T
237	01/12	T/T	50.2	1	C/T	1	C/T
142	30/11	C/C	46.3	1	C/C	2	C/T
430	03/12	T/T	53.3	1	C/T	2	C/T
054	28/11	C/C	45.5	1	C/C	3	C/T
140	30/11	T/T	42.9	1	C/T	3	C/T
272	01/12	C/C	46	1	C/C	3	C/T
268	01/12	T/T	46.6	1	C/T	3	C/T
145	30/11	C/C	45.5	2	C/C	4	C/T
446	03/12	T/T	46.7	2	T/T	5	T/T
289	01/12	C/C	47.9	1	C/C	2	C/T
397	03/12	T/T	48.9	1	T/T	5	T/T
306	01/12	C/C	46	1	C/C	6	C/T
402	03/12	T/T	45.9	1	T/T	5	T/T
412	03/12	C/C	45.1	1	C/C	7	C/T
371	02/12	T/T	42.9	2	C/T	6	C/T

Les brebis grisées n'ont pas été prises en compte dans les analyses pour des raisons décrites dans la section Résultats et discussion.

Afin de limiter les biais expérimentaux dus à d'autres facteurs que celui du génotype *SOCS2*, nous avons pris en compte dans notre choix l'influence génétique, en sélectionnant préférentiellement des duos de demi-sœurs ayant le même père. Cinq duos ont ainsi été initialement constitués (Tableau 8). Les trois autres duos ont été choisis de manière à ce que le groupe intra-génotype soit le plus homogène possible selon les critères de poids, date de mise-bas, et de nombre de fœtus.

II.1.b. Protocole expérimental

Les brebis sont arrivées à l'ENVT le 12 décembre 2017 et ont mis bas entre le 28 décembre 2017 et le 5 janvier 2018. Le protocole général de l'expérimentation animale réalisée dans ce projet de thèse en janvier-février 2018 à l'ENVT est présenté dans la Figure 11.

Les brebis présentées précédemment ont été inoculées dans la nuit du 22 au 23 janvier 2018 à minuit d'un inoculum de 10^3 UFC (Unité Formant Colonie) de *Staphylococcus aureus* (souche SA9A) par voie diathélique, à l'aide d'une sonde stérile à usage unique (volume de 1 mL). Cette inoculation a été réalisée dans des conditions aseptiques (nettoyage et désinfection du trayon à l'aide d'une lingette antiseptique). Le choix de cette souche spécifique de *Staphylococcus aureus* repose sur plusieurs critères : innocuité vis-à-vis de l'opérateur (sensibilité aux antibiotiques), connaissance de la virulence lors d'infections intra-mammaires antérieures chez la brebis (expression clinique franche mais non létale), et représentativité des souches ovines en termes de facteurs de virulence (Deverrière, 2007).

Le choix de l'hémi-mamelle à inoculer a été réalisé grâce à des mesures effectuées sur des échantillons de lait préalablement prélevés de manière aseptique sur les deux hémi-mamelles de l'ensemble des brebis les 4, 9, 11, 16 et 18 janvier 2018 (Figure 11). Deux premières mesures reflètent l'état inflammatoire de chaque hémi-mamelle (CMT - *California Mastitis Test* -, et CCS, du lait), tandis que la troisième reflète l'état infectieux (quantité de bactéries dans le lait). Brièvement, le CMT consiste en l'ajout dans l'échantillon de lait d'un réactif composé d'un détergent et d'un indicateur de pH. Lorsqu'il est mélangé avec le lait, le réactif réagit et forme un gel visqueux, qui peut être noté selon une gamme de couleurs et de viscosité du lait, indiquant ainsi le nombre approximatif de cellules somatiques présentes. Les CMT ont été réalisés avec un test commercial (Leucocyttest, Merial). Les CCS ont été mesurées à l'aide du compteur cellulaire DCC de la marque DeLaval®. La quantification des bactéries a été réalisée par bactériologie conventionnelle par la visualisation et l'estimation du nombre de germes présents dans le lait après culture sur gélose. Grâce à ces analyses, nous avons préférentiellement inoculé l'hémi-mamelle droite des brebis lorsqu'elle était saine et trois brebis ont été inoculées dans l'hémi-mamelle gauche à cause d'une infection de l'hémi-mamelle controlatérale.

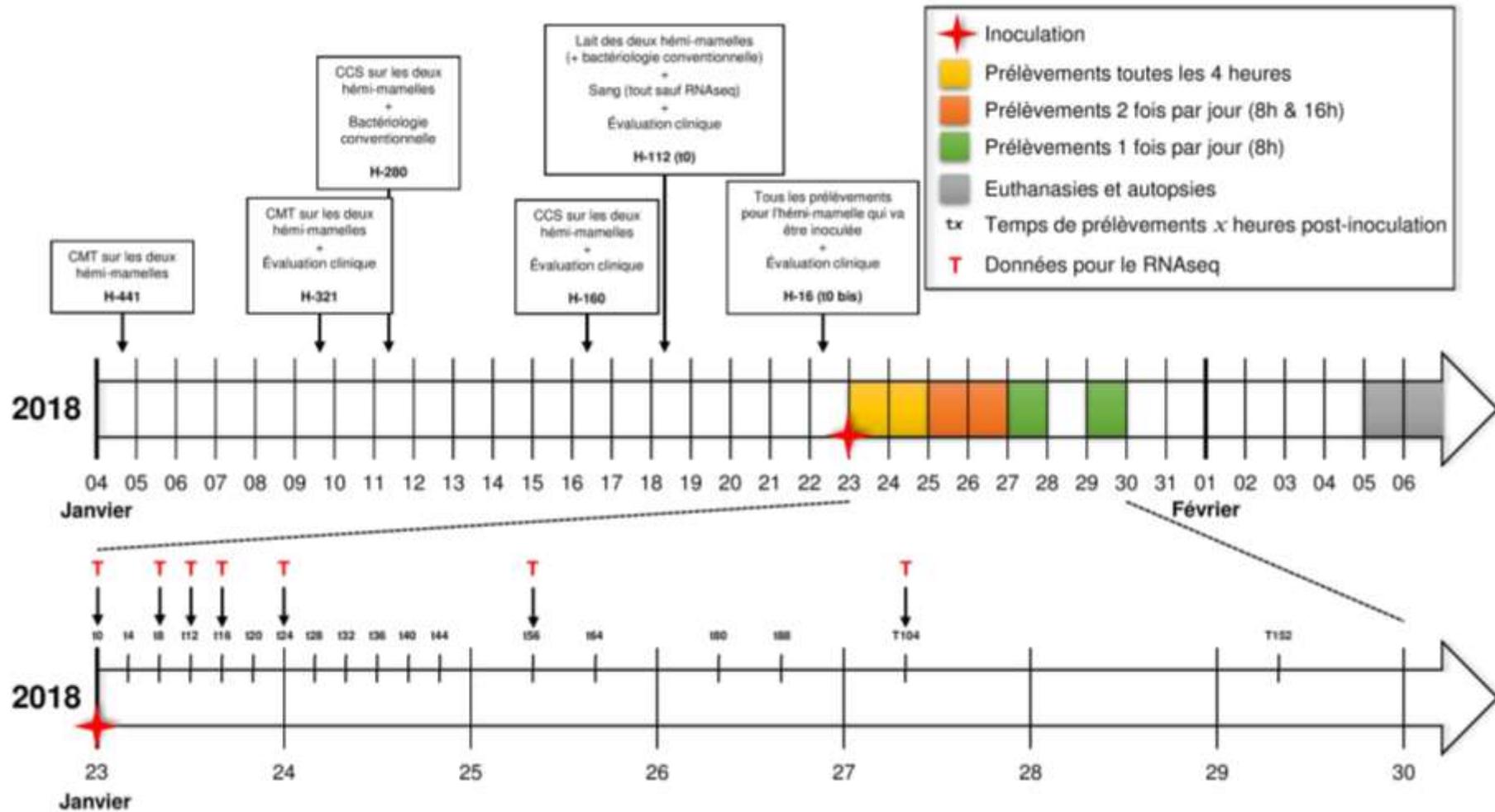


Figure 11 Protocole des infections expérimentales réalisée à l'ENVT en janvier-février 2018
 Abréviations : CMT : California Mastitis Test ; CCS : Comptage de Cellules Somatiques.

Les suivis clinique et la collecte d'échantillons biologiques ont été réalisés pendant 152 heures post-inoculation (Figure 11). La fréquence des mesures et des prélèvements d'échantillons a été de six fois par 24 heures (toutes les quatre heures) durant les deux premiers jours de l'infection expérimentale, puis de deux fois par jour, aux horaires de la traite (8h et 16h) au cours des deux jours suivants, et enfin de une fois par jour (8h) les 27 et 29 janvier 2018 (respectivement 104 et 152 heures post-inoculation, *i.e.* T104 et T152). Les brebis ont ensuite été euthanasiées et autopsiées les 5 et 6 février 2018.

II.2. Modalités de collecte et de traitement des échantillons

L'ensemble des mesures acquises est décrit ci-dessous et résumé dans le Tableau 12 et en Annexe 7.

II.2.a. Suivi clinique

Un examen clinique fin a été réalisé à chaque temps de contrôle par deux étudiantes de l'ENVT (Charlotte Duheron et Laurence Frayssignes) dans le cadre de leur thèse de Doctorat vétérinaire.

II.2.a.i. Modalités de collecte

Deux catégories de données cliniques ont été collectées : des données à l'échelle systémique, afin de quantifier la condition générale de la brebis, et à l'échelle mammaire, afin de visualiser les signes cliniques locaux provoqués par l'infection bactérienne.

A l'échelle systémique, sept phénotypes ont été observés/mesurés. La mesure des fréquences cardiaque, respiratoire, et ruminale a été effectuée, de même qu'une appréciation des états d'appétit, d'abattement et des muqueuses oculaires. L'examen mammaire a généré neuf phénotypes par hémi-mamelle (inoculée et non inoculée) dont sept sont des appréciations : couleur, douleur au toucher, chaleur, induration, gonflement, couleur du lait, et aspect du lait. A ces phénotypes s'ajoutent : un phénotype sur le nombre et la taille des abcès, évalués par palpation mammaire, ainsi qu'un phénotype lié au volume des nœuds lymphatiques mammaires.

En plus de ces mesures cliniques, le relevé de température ruminale à l'aide d'un Thermobolus (Medria Technologies) a également été réalisé à partir du 22 décembre 2017 (avant la mise-bas). Ce dispositif permet d'assurer un suivi continu de la température individuelle du rumen des brebis. Par ailleurs, la production laitière par hémi-mamelle a également pu être notée afin de suivre l'évolution de la production laitière cours de l'infection.

II.2.a.ii. Traitement des échantillons

L'évaluation de la condition clinique à l'échelle systémique a été synthétisée grâce à un score global sur 14 (Tableau 9) intégrant les sept phénotypes décrits précédemment. Pour ce faire, les données brutes des fréquences cardiaque, respiratoire et ruminale ont été transformées en score noté de 0 à 2 afin qu'ils soient dans la même échelle que les autres phénotypes. Pour les fréquences cardiaque et respiratoire, une référence de base (*FB*) a été établie pour chaque animal à partir de trois dates de contrôle recueillies avant l'inoculation (moyenne des valeurs des temps H-321, H-160, H-112, cf. Figure 11). Pour chaque temps de contrôle, les valeurs ont été transformées en pourcentage par rapport aux *FB* et la transformation finale en score a été effectuée selon les critères détaillés dans le Tableau 9. Les valeurs de la fréquence ruminale étant comprises naturellement entre 0 et 2 ruminations par minute, la transformation en score était plus simple. Les appréciations des états d'appétit, d'abattement et de l'aspect des muqueuses oculaires ont été notés sous forme de score compris entre 0 et 2. Le score global sur 14 était alors la somme de l'ensemble de ces sept phénotypes sous forme de score, où plus le score est élevé, plus la dégradation clinique est marquée.

Tableau 9 Critères d'évaluation du score clinique général sur une échelle de 0 à 14

<i>Phénotype</i>	<i>Score</i>		
	<i>0</i>	<i>1</i>	<i>2</i>
<i>Fréquence cardiaque</i>	$x < FB + 20\%$	$FB + 20\% \leq x < FB + 40\%$	$x \geq FB + 40\%$
<i>Fréquence respiratoire</i>	$x < FB + 10\%$	$FB + 10\% \leq x < FB + 30\%$	$x \geq FB + 40\%$
<i>Fréquence ruminale</i>	2	1	0
<i>Température rectale (°C)</i>	$38,5 \leq x \leq 39,5$	$39,5 < x \leq 40$	$x < 38,5$ OU $x > 40$
<i>Appétit</i>	conservé	diminué	absent
<i>Abattement</i>	absent	léger	marqué
<i>Muqueuses oculaires</i>	rosées	congestion légère	congestion marquée

Abréviation : *FB* : Fréquence de Base

A l'échelle mammaire, des scores globaux de l'état clinique des héli-mamelles inoculées et non inoculées ont également été conçus (Tableau 10). Pour ces scores, seul le phénotype du nombre et de la taille des abcès mammaires devait être transformé en score. Les scores globaux sur 12 étaient alors la somme de l'ensemble de ces neuf phénotypes sous forme de score où l'état clinique de l'héli-mamelle est dégradé lorsque le score augmente.

Tableau 10 Critères d'évaluation des scores cliniques mammaires sur une échelle de de 0 à 12

<i>Phénotype</i>	<i>Score</i>			
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>
<i>Couleur</i>	normale	anormale		
<i>Douleur</i>	absence	présence		
<i>Chaleur</i>	absence	présence		
<i>Induration</i>	absence	présence		
<i>Gonflement</i>	absence	présence		
<i>Abcès</i>	absence	1 abcès cm OU < 4 abcès mm		> 1 abcès cm OU ≥ 4 abcès mm
<i>Volume næuds lymphatiques</i>	normal	augmentation modérée	augmentation sévère	
<i>Couleur du lait</i>	blanc	anormal		
<i>Texture du lait</i>	normale	anormale		

Abréviations : cm : centimétrique (taille de 1 centimètre au minimum) ; mm : millimétrique (taille inférieure à 1 centimètre).

Dans les analyses effectuées dans cette thèse, la température a été traitée en tant que donnée numérique. Deux mesures de la température (ruminale et rectale) étaient disponibles mais non complètes en tant que données numériques. En effet, les températures du rumen (x_{rumen}) enregistrées (extraction des données corrigées par les serveurs de l'entreprise Medria Technologies, uniquement aux temps de prélèvements) étaient incomplètes en raison de l'arrêt d'émission, au cours du protocole, de quatre Thermobolus pour des raisons techniques. Les températures rectales ($x_{rectale}$) relevées au cours des examens cliniques étaient principalement relevées dans des intervalles de valeurs ; les valeurs réelles étaient disponibles lorsqu'elles dépassaient les bornes de l'intervalle des valeurs usuelles. J'ai donc choisi de conserver les mesures de la température ruminale (moins de données manquantes, donc plus précis) en complétant les valeurs manquantes grâce aux données rectales. Pour ce faire, les températures

rectales ont été transformées en valeurs numériques en prenant la moyenne des valeurs des intervalles (39°C pour le score 0 et 39,75°C pour le score 1) lorsque la donnée réelle n'était pas notée (la valeur réelle du score 2 était toujours notée). Les valeurs manquantes de la température du rumen d'un individu i à l'instant t ont alors été obtenues en appliquant la formule suivante afin de corriger les différences d'échelle entre les températures du rumen et rectales :

$$x_{rumen,i,t} = \mu_{rumen,t} + \sigma_{rumen,t} \times \frac{x_{rectale,i,t} - \mu_{rectale,t}}{\sigma_{rectale,t}}, \text{ où } \mu_t \text{ et } \sigma_t \text{ sont respectivement la}$$

moyenne et l'écart-type des valeurs pour l'ensemble des 16 brebis.

II.2.b. Prélèvements de lait

Deux types d'échantillons de lait ont été prélevés : un échantillon collecté dans des conditions aseptiques pour les analyses bactériologiques, et un échantillon pour le dénombrement et le typage des cellules somatiques dans le lait.

II.2.b.i. Modalités de collecte

Pour l'hémi-mamelle inoculée, et à tous les temps de contrôle, après élimination manuelle des premiers jets de lait, un prélèvement aseptique de 2 mL environ était réalisé pour l'analyse bactériologique. Cet échantillon était entreposé à 4°C puis rapidement congelé à -20°C pour un traitement différé. Ce prélèvement était suivi d'un prélèvement manuel de 25 mL de lait dans un pot contenant du Bronopol, pour la quantification des cellules dans le lait et la différenciation des types cellulaires. Ce dernier était identifié, entreposé à 4°C et traité sous 72h.

II.2.b.ii. Traitement des échantillons

L'analyse bactériologique du lait des hémi-mamelles inocuées a été réalisée avec une approche développée à l'ENVT et précédemment décrite dans le [Chapitre III](#). (Oget *et al.*, 2019a), une approche basée sur une qPCR (*quantitative Polymerase Chain Reaction* - réaction en chaîne par polymérase quantitative). Le phénotype obtenu à partir de cette approche est une quantification d'ADN de *Staphylococcus aureus* dans le lait et reflète donc la capacité de la brebis à éliminer la bactérie au cours de l'infection.

Les CCS ont été déterminés, avec le compteur cellulaire DCC de la marque DeLaval®. Lorsque la valeur dépassait les 2000 cellules/mL, le lait était dilué au dixième, puis une nouvelle mesure était réalisée afin que la concentration soit dans l'intervalle de linéarité de la mesure. La dilution était répétée jusqu'à ce que la valeur obtenue soit exploitable (< 2000 cellules/mL). Cette mesure a été effectuée au fur et à mesure du protocole, ce qui a permis de réaliser un suivi fin en temps réel de l'inflammation de l'hémi-mamelle inoculée. Une transformation logarithmique

du comptage en score de cellules somatiques, couramment utilisée pour approcher la variable CCS de la normalité (Ali & Shook, 1980), a été appliquée à l'aide de la formule suivante :

$$SCS = \log_2\left(\frac{SCC}{100}\right) + 3.$$

Le typage des cellules du lait a été réalisé par cytométrie en flux sur le culot cellulaire de l'échantillon utilisé pour l'analyse des CCS, après centrifugation du lait et élimination de la crème et du lactosérum. L'acquisition et l'analyse des données de cytométrie ont été effectuées respectivement avec le cytomètre MACSQuant® (Miltenyi Biotec) et le logiciel FlowJo®. Les différentes populations leucocytaires ont ainsi pu être caractérisées et quantifiées (variables exprimées en nombre de cellules par mL de lait). Les populations quantifiées sont les granulocytes neutrophiles vivants et morts, les monocytes et macrophages, et les lymphocytes.

II.2.c. Prélèvements de sang

Deux types de prélèvements de sang ont été effectués : le premier (anticoagulant EDTA - acide ÉthylèneDiamineTétraAcétique) était destiné aux analyses d'expression génique, à la quantification des cellules sanguines, et au dosage des protéines inflammatoires, et le second (anticoagulant héparine de lithium), au dosage des cytokines.

II.2.c.i. Modalités de collecte

Les prélèvements de sang ont été réalisés à chaque temps de contrôle par des personnes compétentes par ponction jugulaire à l'aide d'un porte aiguille associé à un vacutainer.

Le premier prélèvement (EDTA), destiné aux analyses d'expression génique, à la numération et formule sanguines, au typage fin des différentes populations cellulaires présentes dans le sang, et au dosage de protéines inflammatoires, était dans un premier temps entreposé à 4°C sur la glace, puis rapidement aliquoté en quatre exemplaires avec du tampon de lyse DL, le Thiocyanate de Guanidine (quatre fois : 400 µL de sang total + 400 µL de tampon DL). Ces plaques étaient ensuite refroidies à -20°C pendant une heure puis stockées à -80°C pour une extraction différée des ARN. Le sang total restant du tube EDTA était ensuite centrifugé et le plasma aliquoté en double exemplaire sur deux plaques différentes (deux fois 500 µL) ainsi que dans un tube de volume de 1 mL. Ces échantillons ont ensuite été conservés à -80°C pour un traitement différé (dosage des protéines inflammatoires).

Le second prélèvement (héparine) était destiné au dosage des cytokines. Le sang était centrifugé et le plasma aliquoté en double exemplaire sur deux plaques différentes (deux fois 500 µL) ainsi

que dans un tube de volume de 1 mL. Ces échantillons ont ensuite été conservés à -80°C pour un traitement différé (dosage des cytokines).

II.2.c.ii. Traitement des échantillons

Expression génique

Les plaques congelées avec le tampon de lyse DL ont été traitées ultérieurement par Julien Sarry (GenPhySE, INRA). Les ARN ont été extraits uniquement pour sept dates choisies d'après les résultats préliminaires disponibles (CCS) et justifiés dans la section Résultats et discussion : T0, T8, T12, T16, T24, T56 et T104, comme indiqué sur la Figure 11. Pour des raisons techniques, l'ARN de certains échantillons n'a pas pu être extrait et un temps supplémentaire a été ajouté (T88) en complément du dernier temps tardif (T104) (Tableau 11). Ces deux temps ont été regroupés pour les analyses (réponse immunitaire tardive).

Tableau 11 Effectif disponible par temps pour les analyses RNA-seq de l'infection expérimentale

<i>Temps</i>	<i>Données disponibles</i>
<i>T0</i>	14 brebis
<i>T8</i>	14 brebis
<i>T12</i>	12 brebis (402 et 446 manquantes)
<i>T16</i>	14 brebis
<i>T24</i>	13 brebis (054 manquante)
<i>T56</i>	14 brebis
<i>T104</i>	T88 : 054, 289, 402 et 446
<i>(T88 et T104 regroupés)</i>	T104 : 140, 142, 145, 237, 268, 272, 306, 397, 402, 472, et 430

Pour le dernier temps, nous avons regroupé T88 et T104 en vérifiant au préalable la cohérence des comptages sur l'individu doublé (402) qui a été par ailleurs conservé en un seul exemplaire (T104).

L'extraction des ARN a été réalisée à l'aide du kit commercial Nucleospin® 96 RNA Blood (Macherey-Nagel). La quantification et la vérification de la qualité des ARN extraits ont été effectuées respectivement par mesure de l'absorbance (NanoDrop ND-1000®) et à l'aide d'un Fragment analyzer®. Les bibliothèques ont été préparées à partir des ARN de haute qualité grâce à un kit commercial de préparation d'échantillons Illumina TruSeq® RNA. Ces échantillons ont été amplifiés par PCR, quantifiés par qPCR, puis séquencés par un séquenceur Illumina

NovaSeq 6000® à la plateforme génomique de Genotoul (Castanet-Tolosan, France). Les séquences étaient alors disponibles en format FASTQ.

Phénotypes

L'haptoglobine, une protéine inflammatoire plasmatique a été dosée à partir des aliquotes congelées de plasma prélevé sur prélèvement EDTA à l'aide d'un test ELISA (BioX Diagnostics).

La numération, la formule et le typage fin des cellules sanguines ont été effectués immédiatement après le prélèvement par cytométrie en flux sur les aliquotes du prélèvement EDTA. Un exemple d'analyse des données de cytométrie (lymphocytes) est fourni en Annexe 8. Les populations cellulaires caractérisées et quantifiées sont les suivantes :

- PBMC (*Peripheral Blood Mononuclear Cell* - cellules mononuclées périphériques sanguines)
 - Leucocytes totaux
 - Monocytes totaux
 - Monocytes conventionnels, ou classiques
 - Monocytes inflammatoires, ou non classiques
 - Monocytes intermédiaires
 - Neutrophiles
 - Lymphocytes totaux
 - Lymphocytes B (CD21)
 - Lymphocytes (CD2)
 - Lymphocytes T CD2 CD4 (auxiliaires)
 - Lymphocytes T CD2 CD8 (cytotoxiques)
 - Lymphocytes CD2 non CD4, non CD8
 - Lymphocytes non B, non T
 - Lymphocytes NK (CD16)
 - Lymphocytes totaux sans les NK
 - Ratio entre neutrophiles et monocytes inflammatoires
 - Ratio entre lymphocytes T CD4 et lymphocytes T CD8

A partir des échantillons de plasma du tube Héparine, des cytokines ont été dosées : les interleukines IL-1a, IL-1b, IL-1ra, IL-2, IL-4, IL-6, IL-8, IL-10, IL-17a, les interférons IP-10, IFN- γ , les chimiokines CCL-2, CCL-3, CCL-4, et le facteur de nécrose tumorale TFN- α , à l'aide d'un test Milliplex® Custom bovine/ovine cytokines (Merck Millipore).

II.2.d. Mesures d'autopsies

A la fin du protocole, les brebis ont été euthanasiées puis autopsiées. La carcasse et les organes internes ont été examinés (pesées, et mesures pour les os longs). Dix-sept phénotypes ont ainsi pu être collectés :

- Deux phénotypes concernant les abcès : quantité (variable qualitative, non analysée ultérieurement) et taille.
- Onze phénotypes de poids : vif, carcasse, cœur, foie, poumon, rate, reins, hémimamelles inoculée et non inoculée, nœuds lymphatiques mammaires de l'hémi-mamelle inoculée et non inoculée,
- Quatre dimensions des os (grand axe) : fémur, humérus, radius, et tibia.

II.3. Traitement bioinformatique et analyses statistiques

II.3.a. Analyses des phénotypes

Les analyses préliminaires des phénotypes générés au cours de l'expérience sont résumées dans le Tableau 12. Toutes les analyses statistiques ont été réalisées avec le logiciel R version 3.5.0 (R Development Core Team, 2005).

Le but étant de déterminer les effets du génotype *SOCS2* sur les différents phénotypes, nous avons appliqué des modèles linéaires mixtes (fonction `lme()` du package R *nlme* version 3.1.137 (Pinheiro *et al.*, 2019)) sur les mesures pour lesquelles nous avons des données répétées. Ces modèles incluaient un effet aléatoire animal (la disponibilité des temps de répétition des données est fournie en Annexe 7), ainsi que les effets fixes suivants : génotype *SOCS2*, temps, et interaction génotype *SOCS2* et temps. Pour certains phénotypes indiqués dans le Tableau 12 (†), dont les données étaient légèrement dissymétriques, nous avons effectué une transformation logarithmique sur les données brutes ($\log(x + 1)$) pour s'approcher au mieux de la distribution normale.

Les effets des génotypes intra-temps (comparaisons entre les niveaux de génotypes deux à deux intra-temps, c'est-à-dire tests *post-hoc*) ont ensuite été évalués, avec une correction pour les tests multiples à l'aide de la procédure de Bonferroni (fonction `lsmeans()` du package R *lsmeans* v2.23 (Lenth, 2016)). Concernant les phénotypes pour lesquels une transformation logarithmique a été effectuée, les moyennes corrigées par le modèle (`lsmeans`) ont été transformées dans l'échelle du phénotype (paramètre de la fonction `lsmeans()` : `type = "response"`).

Tableau 12 Récapitulatif des données recueillies et générées au cours de l'infection expérimentale et des analyses des variables déduites

<i>Nature des Données</i>	<i>Mesures</i>	<i>Variables finales</i>	<i>Analyses</i>
<i>DONNEES CLINIQUES</i> 28 variables	Monitoring continu de la température Examen clinique	température ruminale imputée scores : systémique [†] et mammaires inoc. [†] et non inoc. [†] ; fréquences : cardiaque, respiratoire et ruminale ; scores appétit, abattement et muqueuses ; scores couleur, douleur, chaleur, induration, gonflement, abcès, nœuds lymphatiques, couleur lait et texture lait (inoc. et non inoc.)	Modèles mixtes (sauf scores, à l'exception des scores résumés)
<i>DONNEES LAIT</i> 9 variables	Bactériologie Cellules somatiques Production laitière Typage cellules	bactériologie par qPCR (inoc.) [†] SCS (inoc.) productions totale et par hémi mamelles (inoc. et non inoc.) neutrophiles vivants [†] et morts [†] , monocytes et macrophages [†] , et lymphocytes (inoc.) [†]	Modèles mixtes
<i>DONNEES SANG</i> 34 variables (+ RNA-seq)	Expression ARN Dosage protéine inflammatoire Typage cellules	comptages de gènes haptoglobine [†] leucocytes [†] , monocytes [†] , neutrophiles [†] , lymphocytes totaux [†] , B [†] , T [†] , et non B et non T [†]	Modèles linéaires généralisés Modèles mixtes

	Typage détaillé	monocytes : conventionnels [†] , inflammatoires [†] et intermédiaires [†] ; lymphocytes : NK [†] , sans NK [†] ; ratio entre neutrophiles et monocytes inflammatoires [†] , et PBMC [†]	
	Typage lymphocytes T	lymphocytes T : CD4+ [†] , CD8+ [†] , CD4- et CD8- [†] , ratio entre CD4+ et CD8+ [†]	
	Dosage cytokines	interleukines : IL1a [†] , IL1b [†] , IL1ra [†] , IL2 [†] , IL4 [†] , IL6 [†] , IL8 [†] , IL10 [†] , IL17a [†] ; interférons : IP10 [†] , IFN- γ [†] ; chimiokines : CCL2 [†] , CCL3 [†] , CCL4 [†] ; facteur de nécrose tumorale : TNF- α [†]	
<i>DONNEES AUTOPSIES 16 variables</i>	Pesées et mesures	taille des abcès [¥] ; pesées : vif, carcasse, cœur, foie, poumon, rate, reins, hémi-mamelles (inoc. [¥] et non inoc.), et nœuds lymphatiques (inoc. et non inoc. [¥]) ; mesures : fémur, humérus, radius, et tibia	Comparaisons de moyennes

Abréviation : inoc. : inoculée ; SCS : Score de Cellules Somatiques. Le symbole [†] signifie que les données brutes ont subi une transformation logarithmique pour s'approcher au mieux de la normalité dans le modèle mixte. Pour les données d'autopsies, le symbole [¥] signifie que la normalité n'était pas respectée et que le test de comparaison de moyenne effectué est non paramétrique (Wilcoxon-Mann-Whitney) à l'inverse des autres données (Student).

Les données provenant des autopsies n'étant pas répétées (disponibles uniquement à l'abattage), des tests de comparaison de moyenne ont été réalisés pour déterminer l'effet de la mutation *SOCS2* sur ces phénotypes. Nous avons utilisé le test de Student pour les variables qui suivaient la normalité, et le test de Wilcoxon-Mann-Whitney pour les autres (indiqué dans le Tableau 12).

II.3.b. Analyse des données d'expression génique par RNA-seq

II.3.b.i. Traitement bioinformatique

Les séquences ont été produites par la plateforme génomique de Genotoul et livrées au format FASTQ. Le traitement bioinformatique de ces fichiers a été réalisé avec le soutien du bioinformaticien Cédric Cabau (GenPhySE, INRA), et est présenté en détails en Annexe 9.

Le génome de référence sur lequel nous avons choisi d'aligner les séquences est l'assemblage ovin le plus récent (https://www.ncbi.nlm.nih.gov/assembly/GCF_002742125.1/), Oar_rambouillet_v1.0.

Les lectures ont été nettoyées et leur qualité vérifiée, à l'aide respectivement des logiciels Cutadapt v2.1 (Martin, 2011) et fastQC v0.11.7 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Un premier alignement des lectures sur le génome de référence a été effectué, suivi d'une quantification des transcrits d'après le transcriptome de référence. Ce modèle de référence a ensuite été étendu afin de mieux rendre compte de la diversité des transcrits observés. Un second alignement suivi d'une quantification des transcrits décrits par ce nouveau modèle a alors été réalisé. Ce modèle étendu a été constitué à partir de 28 individus (quatre par temps), choisis parmi les fichiers d'alignements (BAM) les plus lourds par temps, et avec équilibrage des effectifs pour les deux groupes de génotypes *SOCS2* (deux T/T et deux C/C par temps) et pour la représentation de l'ensemble des individus (tous les individus représentés au moins une fois et au maximum trois fois). Les alignements, les quantifications des transcrits, et la constitution du modèle étendu ont été réalisés respectivement à l'aide des logiciels STAR v2.6 (Dobin *et al.*, 2013), RSEM v1.3.0 (Li & Dewey, 2011), et Cufflinks v2.2.1 (Trapnell *et al.*, 2010).

II.3.b.ii. Analyses statistiques et fonctionnelles

Les analyses statistiques sur les comptages bruts d'expression de gènes ont été réalisées à l'aide de modèles linéaires généralisés avec le package R *EdgeR* v 3.20.9 (Robinson *et al.*, 2010). Pour toutes les analyses, les comptages bruts ont été filtrés d'après les recommandations du manuel d'utilisation du package R *EdgeR* en conservant uniquement les gènes comptabilisant

au moins 10 lectures dans un nombre de librairies correspondant au nombre minimal d'individus par groupe. En effet, les tailles des librairies étaient, dans notre jeu de données, en moyenne de 21 428 655 lectures (comprises entre 9 998 339 et 34 21 007), c'est-à-dire qu'en conservant les gènes dont la somme des comptages dans l'ensemble des 95 librairies était de 0,5 comptages par million, nous conservions en moyenne 10,7 lectures dans chaque groupe.

Deux études ont été conduites :

- La première a pour objectif d'identifier les gènes différentiellement exprimés au cours du temps de l'expérience afin de caractériser le profil d'expression génique de la réaction immunitaire en réponse à une infection bactérienne. Cette étude a été effectuée dans un premier temps sans prendre en compte le génotype *SOCS2*. Dans un second temps, l'effet du génotype a été intégré (test des contrastes intra-génotype C/C, et intra-génotype T/T, séparément). Ces deux approches permettent la comparaison des trois listes de gènes différentiels à chaque temps (intra-génotype et tous génotypes *SOCS2*). Pour la première approche (intra-temps), la taille minimale des groupes (cf. Tableau 11) était de 12 brebis, et pour la seconde (intra-temps et intra-génotype), de 5 brebis.
- La seconde étude a pour objectif d'identifier les gènes différentiellement exprimés entre les deux groupes d'individus homozygotes mutants T/T et sauvages C/C au gène *SOCS2* afin de caractériser les modifications transcriptionnelles en lien avec la mutation génétique. La taille minimale des groupes était ici la même que précédemment (seconde approche intra-temps et intra-génotype), c'est-à-dire 5.

Après filtrage, les comptages ont été normalisés avec la méthode TMM (*Trimmed Mean of M-values* - moyenne ajustée des valeurs M) (Robinson & Oshlack, 2010) implémentée dans le package *EdgeR*. Des modèles linéaires généralisés ont ensuite été appliqués avec uniquement un effet fixe Temps pour la première approche de la première étude (réponse sanguine sans considération du génotype *SOCS2*), et avec trois effets fixes pour les autres approches : Temps, Génotype et Interaction Temps/Génotype, à l'aide de la fonction `glmFit()`. Les comptages des gènes ont été testés pour les contrastes d'intérêt à l'aide des fonctions `glmLRT()` et `decideTestsDGE()`, avec une correction pour les tests multiples à l'aide de la procédure de Benjamini et Hochberg (seuil de 5%) (Benjamini & Hochberg, 1995).

Des analyses fonctionnelles ont été réalisées à l'aide du logiciel IPA® (Ingenuity Pathway Analysis) (Krämer *et al.*, 2014) sur les listes des gènes différentiellement exprimés.

L'assemblage du génome ovin choisi pour l'alignement étant récent, une annotation manuelle a parfois été réalisée à partir des gènes orthologues chez la souris et chez l'homme proposés par le site internet NCBI.

II.3.c. Intégration de l'ensemble des données

L'intégration de l'ensemble des données phénotypiques et d'expression génique de l'expérience a été réalisée à l'aide du package R *MixOmics* v6.6.1 (Lê Cao *et al.*, 2009). Cette intégration est une approche complémentaire et de validation des approches précédemment utilisées (modèles et tests statistiques). En effet, il s'agit de méthodes de classification qui permettent de mettre en valeur les variables qui expliquent le plus la variabilité des données collectées au cours de ce protocole, sans ou avec une supervision par les variables qualitatives discriminantes d'intérêt, *i.e.* temps et génotype *SOCS2*.

Deux jeux de données ont été étudiés.

Le premier permet la mise en lumière, à l'échelle du protocole, des variables les plus explicatives des différences entre les individus. Il comprend l'ensemble des données disponibles pour les sept temps du transcriptome (95 échantillons), pour un total de 19 866 variables (trois variables qualitatives incluses : génotype *SOCS2*, temps et animal). Trois blocs de variables quantitatives ont été constitués :

- données cliniques partielles (N = 6 variables) : température ruminale imputée, fréquences cardiaque et respiratoire, et trois scores globaux,
- données phénotypiques (N = 25 variables) : bactériologie et cellules somatiques dans le lait, dosage de la protéine inflammatoire (haptoglobine), typage des cellules sanguines, et dosage des cytokines dans le sang,
- données d'expression génique par RNA-seq (N = 19 832 variables) : gènes exprimés, filtrés et normalisés.

Le second est constitué des données disponibles au temps T56. Ce jeu de données comprend un total de 19 884 variables (données qualitatives incluses) pour 14 échantillons. Quatre blocs de variables ont été créés (toutes les variables sont disponibles pour ce temps, à l'exception des variables d'autopsies) : données cliniques (N = 5 variables, la fréquence respiratoire ayant été écartée au vu des résultats présentés dans la partie III.1.b.ii), données laitières (N = 9 variables), données sanguines (N = 34 variables), et données de comptages (N = 19 832 variables).

Pour chacun de ces jeux de données, les analyses suivantes ont été appliquées :

- une analyse en composantes principales PCA (*Principal Component Analysis*) pour chaque bloc individuellement,
- une analyses discriminante dirigée PLS-DA (*Partial Least Square Discriminant Analysis*) pour chaque bloc, pour les facteurs discriminants génotype et temps de contrôle pour le premier jeu de données (7 temps de mesure) et uniquement génotype pour les deux autres (T56),
- une analyse discriminante dirigée multi-bloc PLS-DA permettant l'intégration de l'ensemble des blocs de manière dirigée.

Chaque analyse a été réalisée sans et avec l'option « *sparse* » pour une sélection des variables (N = 6 à 10) les plus discriminantes intra-bloc. De plus, pour le premier jeu de données, chaque analyse a été également réalisée sans et avec l'option « *multilevel* » afin de diminuer la variabilité intra-individuelle (effet de la répétition de l'animal au cours de l'infection expérimentale), à l'exception des analyses discriminantes pour le facteur génotype (problème de convergence en raison probablement de la redondance de l'influence génétique de l'animal et du facteur génotype *SOCS2*).

III. Résultats et discussion

III.1. Analyses des phénotypes

III.1.a. Choix des brebis et des temps analysés d'après le profilage des SCS

Nous avons dans un premier temps analysé les SCS (Figure 12). Le SCS dans le lait augmente très fortement à partir de T8, traduisant un premier recrutement de cellules immunitaires dans la mamelle et le lait en réaction à l'infection bactérienne. Ce nombre atteint une valeur maximale à T24, puis diminue pour atteindre une valeur minimale à T56. Cette valeur augmente une seconde fois à partir de T64 (deuxième pic entre T88 et T104). A partir de T64, les réponses sont plus hétérogènes mais l'on observe une tendance à la baisse du SCS au dernier temps de contrôle (T152).

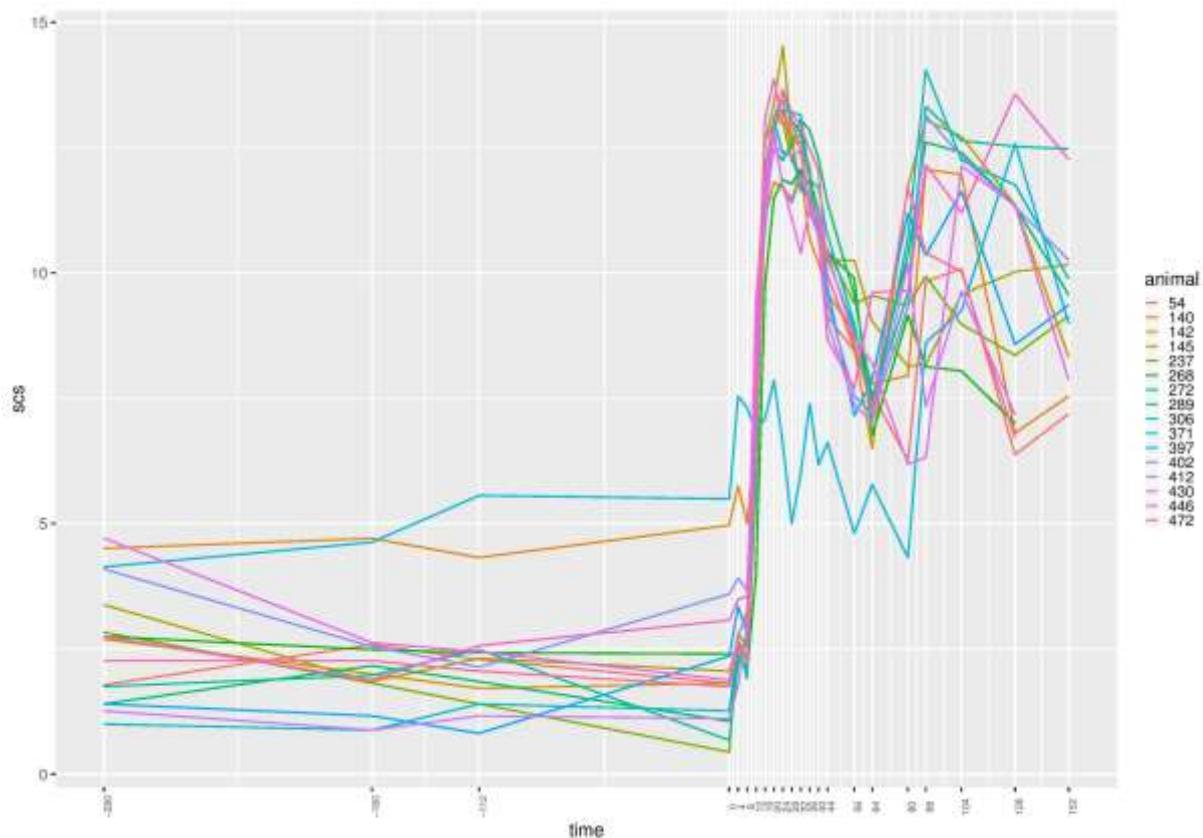


Figure 12 Evolution des scores de cellules somatiques dans le lait de l'ensemble des seize brebis infectées au cours de l'infection expérimentale

Ces résultats ont mis en évidence un individu qui ne suivait pas le même profil de réaction immunitaire que les autres, la brebis 371, génotypée T/T (Figure 12, en bleu). En effet, cette brebis, bien qu'infectée au même moment que les autres, n'a pas eu de recrutement cellulaire en réaction à l'infection. D'après les résultats des analyses de bactériologie effectuées, ceci

s'explique par la présence d'une infection par *Staphylococcus aureus* dans l'hémi-mamelle inoculée, empêchant ainsi l'installation de la souche inoculée. Pour cette raison, la brebis 371 a été écartée des analyses ultérieures. Afin d'équilibrer les effectifs des deux groupes, nous avons également écarté une brebis génotypée C/C des analyses ultérieures, la brebis 412 choisie car elle ne possédait pas de demi-sœur parmi les T/T.

Après élimination de ces deux brebis, nous nous sommes basés sur ce profilage des SCS au cours de l'infection, pour choisir les temps de collecte à retenir pour les analyses RNA-seq. Les temps conservés pour l'analyse du transcriptome sanguin sont les temps 0, 8, 12, 16, 24, 56 et 104. Ces temps permettent de suivre précisément le premier pic de recrutement cellulaire (T0 à T24), puis la phase de dépression entre les deux pics (T56), et enfin la seconde vague de réponse plus tardive (T104).

III.1.b. Des phénotypes révélateurs d'un contrôle de la réaction immunitaire différent entre les sauvages et les mutants

L'objectif des analyses est de décrire des phénomènes biologiques sur la base de l'étude des données et de l'utilisation de la statistique pour confirmer les différences observées. Les résultats de l'ensemble des modèles mixtes effectués sur les variables indiquées précédemment sont présentés en Annexe 10. Nous parlerons dans cette partie de valeurs significatives pour des p-valeurs inférieures à 0,05, et de tendances pour des p-valeurs proches de la significativité, inférieures à 0,1 (seuil suggestif proposé au vu de l'effectif de l'expérimentation : sept individus par groupe et par temps).

III.1.b.i. Des phénotypes impactés par le déroulé de l'infection

Le premier résultat obtenu (non montré) d'après l'ensemble des 49 modèles mixtes effectués est le suivant : l'effet fixe temps est significatif pour la très grande majorité des variables (94%).

Trois phénotypes font exception. En effet, l'effet temps n'est pas significatif pour la variable ratio entre neutrophiles et monocytes inflammatoires. Ceci est probablement dû à la nature du phénotype (rapport), c'est-à-dire que les quantités varient au cours du temps mais pas les proportions. Il n'est pas significatif non plus, bien que très proche (p-valeur = 0,0639), pour la quantité de lymphocytes T CD4- et CD8- dans le sang. Ceci est potentiellement dû à un manque de puissance statistique (uniquement quatre temps disponibles pour ces variables de typage précis des lymphocytes T, ce qui est le plus faible nombre utilisé parmi l'ensemble des modèles mixtes). Enfin, la quantité d'interleukine 1 β (IL1b) dans le sang qui traduit l'action pro-

inflammatoire montre une légère variation à T12 mais reste stable les autres temps, rendant ainsi l'effet temps non significatif à l'échelle du protocole dans le modèle mixte.

Les autres phénotypes étudiés sont impactés par l'infection bactérienne (évolution au cours du temps post-inoculation).

III.1.b.ii. Une réaction clinique accrue chez les T/T

Plusieurs mesures relatives à l'état clinique des brebis sont significativement différentes en fonction du génotype *SOCS2*. La température ruminale (Figure 13) est plus élevée chez les T/T entre les deux pics de recrutement cellulaire dans la mamelle, avec une significativité à T64 (p-valeur = 0,0181) et une tendance à partir de T40 (p-valeurs égales à 0,0504, 0,0673 et 0,0667 respectivement pour T40, T44 et T56). Ce résultat traduit une hyperthermie qui reste marquée chez les brebis T/T alors qu'elle diminue nettement chez les brebis C/C.

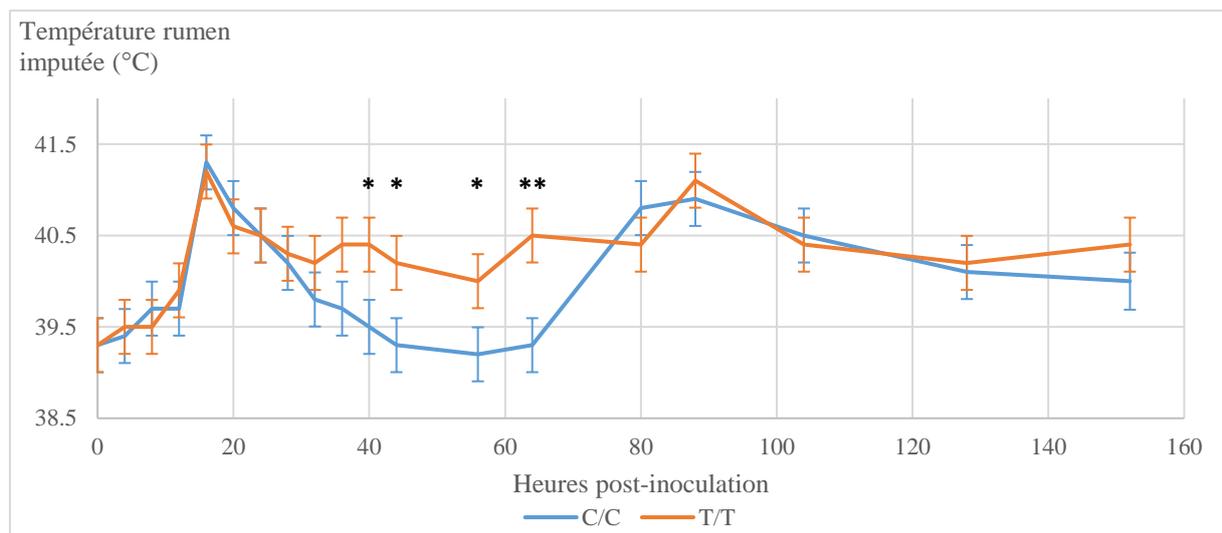


Figure 13 Évolution des lsmeans intra-génotype *SOCS2* de la température du rumen imputée au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10)

La fréquence respiratoire est plus élevée chez les C/C à T56 (tendance, p-valeur = 0,0766, Annexe 10). Pour ce résultat, nous ne proposons pas de claire explication. En revanche, il est à prendre en compte que ce phénotype est une mesure probablement peu précise pour deux raisons. La première est que la durée de mesure est courte (15 secondes pour une approximation d'une minute). La seconde est que la fréquence respiratoire varie de manière instantanée au cours d'une activité et que les individus sont mesurés juste avant la traite (avec distribution de concentré). Ces conditions ne sont alors peut-être pas optimales pour ce type de mesure.

Le score clinique de l'hémi-mamelle inoculée (Figure 14) est plus élevé chez les T/T, de manière significative à T64 (p-valeur = 0,0133). On observe déjà cette tendance à T56 (p-valeur = 0,0869) et par ailleurs, cette tendance se retrouverait sur l'ensemble du protocole puisque l'effet fixe *SOCS2* global est proche de la significativité (p-valeur = 0,0715). La réponse clinique mammaire est donc plus marquée (rougeur, chaleurs induration, etc.) chez les brebis T/T et plus particulièrement entre les deux pics de recrutement cellulaire dans la mamelle.

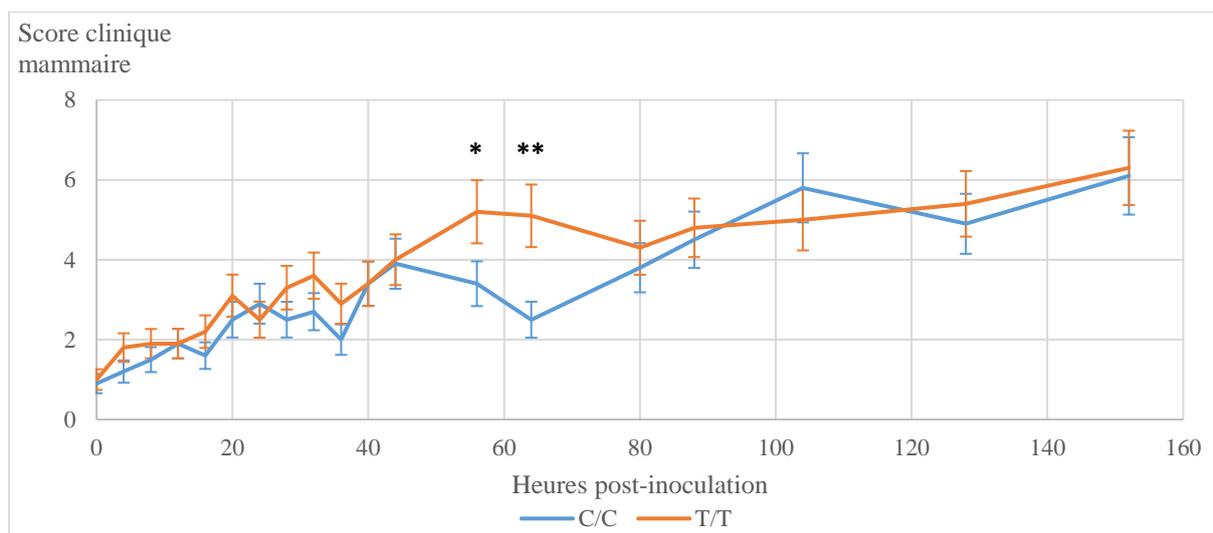


Figure 14 Évolution des lsmeans intra-génotype *SOCS2* du score clinique mammaire de l'hémi-mamelle inoculée au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10)

Des différences cliniques (température ruminale et score clinique mammaire) sont donc observées entre les deux pics de recrutement des cellules immunitaires dans le lait. Ceci semble indiquer une différence de régulation de la réaction immunitaire entre les deux génotypes, avec une dégradation clinique accrue chez les animaux porteurs de la mutation.

III.1.b.iii. Moins de cellules immunitaires recrutées dans le lait chez les T/T

Nous allons maintenant nous intéresser aux variables mesurées dans le lait.

De manière surprenante, la quantité de bactéries dans le lait n'est pas différente entre les C/C et les T/T (Annexe 10). L'association de ce même phénotype à la mutation *SOCS2* avait pourtant été montrée dans le [Chapitre III](#). (Oget *et al.*, 2019a). Ceci est probablement dû à la grande variabilité du phénotype intra-génotype (par les T/T principalement).

La quantité de lait produite varie en réponse à l'infection mais ne semble pas être impactée par la mutation *SOCS2* car aucun effet n'est montré (Annexe 10).

Concernant le profilage des SCS au cours de l'infection, bien qu'aucune différence entre génotypes ne soit significative, des tendances sont notables. Graphiquement (Figure 15), on observe qu'à T0 (mamelle saine), les brebis T/T ont en moyenne un SCS dans le lait plus élevé que les C/C. Ceci est ce qui est observé à plus grande échelle par les GWAS (Rupp *et al.*, 2015). Cette observation s'inverse au moment des deux pics d'inflammation mammaire (SCS plus élevé chez les C/C). Statistiquement (Annexe 10), les différences observées ne sont pas significatives dans notre protocole, mais la tendance est cependant confirmée à T24 (p-valeur = 0,0884) et à T88 (p-valeur = 0,0689). Il y aurait donc plus de cellules immunitaires recrutées chez les C/C. On peut ici faire l'hypothèse que ce recrutement précoce accru de cellules immunitaires dans la mamelle chez les C/C permet une meilleure régulation de l'inflammation (score clinique mammaire diminué à T64, cf. Figure 14).

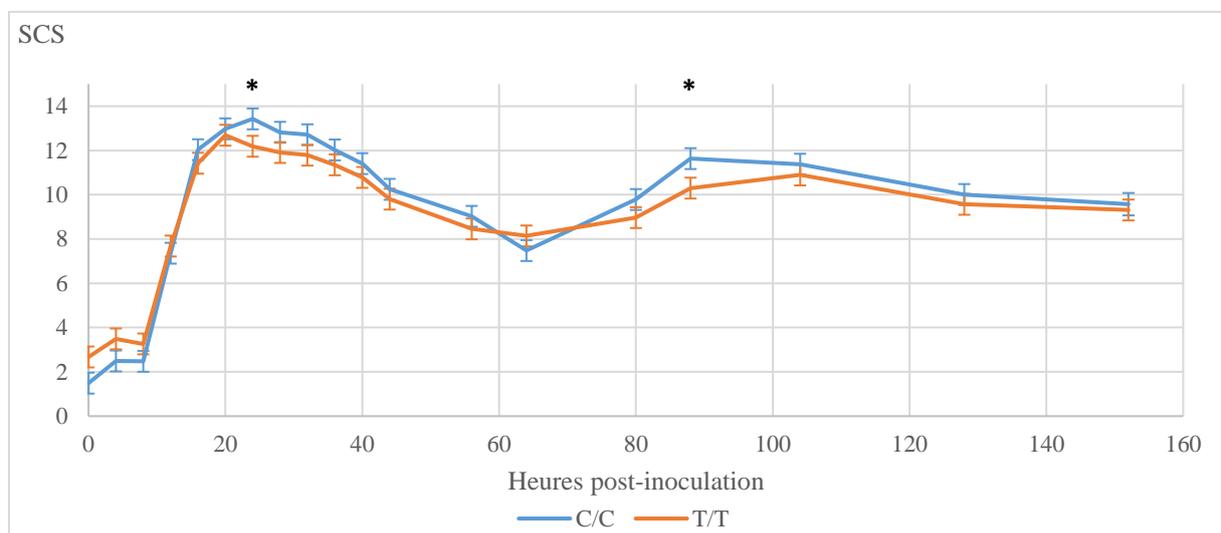


Figure 15 Évolution des lsmeans intra-génotype SOCS2 du SCS du lait de l'hémi-mamelle inoculée au cours de l'infection expérimentale (* : p-valeur < 0,10)

Concernant la composition des cellules somatiques du lait, le premier pic de recrutement de cellules (seul le premier pic est observable car les données ne sont pas disponibles pour le deuxième pic) est observé pour l'ensemble des types cellulaires : neutrophiles, monocytes et macrophages, et lymphocytes (Annexe 10). Les différences observées de quantités des trois ensembles de types cellulaires entre génotypes ne sont pas significatives dans notre protocole. Cependant, on observe une tendance (p-valeur = 0,0633) en faveur d'une quantité de lymphocytes supérieure chez les C/C à T24 (Figure 16). Cette différence à T24 est également présente mais non significative pour les autres types cellulaires, dans l'ordre décroissant : monocytes et macrophages (p-valeur = 0,1721), et neutrophiles vivants (p-valeur = 0,5397), et morts (p-valeur = 0,5418). Les populations cellulaires qui sembleraient donc être moins présentes chez les T/T par rapport aux CC au moment des pics de recrutement cellulaires dans

la mamelle sont donc principalement les lymphocytes, et dans une moindre mesure les monocytes et macrophages.

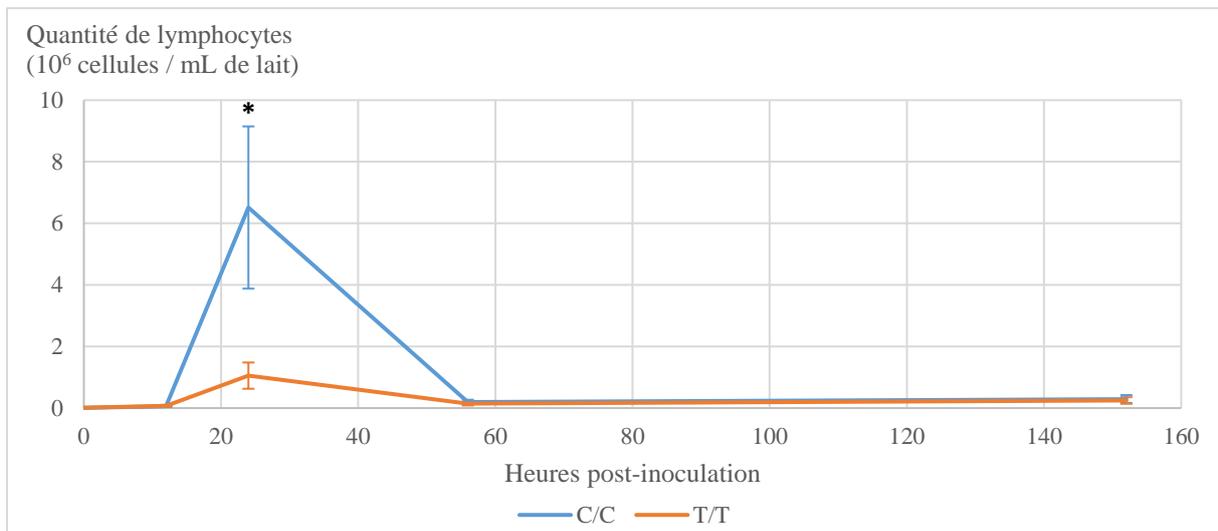


Figure 16 Évolution des lsmeans intra-génotype SOCS2 de la quantité de lymphocytes dans le lait de l'hémi-mamelle inoculée au cours de l'infection expérimentale (* : p-valeur < 0,10)

Les cellules présentes dans le lait sont le reflet des cellules du sang, nous allons voir à présent de manière plus détaillée la composition des cellules sanguines.

III.1.b.iv. Plus de lymphocytes T dans le sang des T/T malgré une composition sanguine proportionnellement inchangée

L'évolution de la concentration moyenne intra-génotype SOCS2 des différents types de cellules sanguines est présentée en Annexe 11. Le nombre de cellules dans le sang évolue à l'inverse du nombre de cellules dans le lait, c'est-à-dire qu'on observe une diminution du nombre de cellules entre T16 et T24, puis à entre T80 et T104. Cela peut être interprété comme un recrutement très marqué de cellules vers la mamelle qui n'est pas compensé par la repopulation à partir de cellules de la moelle osseuse.

Les différences de concentrations entre les C/C et les T/T ne sont pas significatives pour les grands types cellulaires (monocytes, lymphocytes et neutrophiles, cf. Annexe 10). En revanche, on observe significativement moins de lymphocytes T chez les C/C à T20 (p-valeur = 0,0464), et cette tendance est également visible à T16 (p-valeur = 0,0699) (Figure 17). Cela signifierait que plus de lymphocytes T passent dans la mamelle au moment du pic de recrutement de cellules immunitaires chez les C/C. Or, les lymphocytes font partie de l'immunité adaptative (Sordillo, 2018).

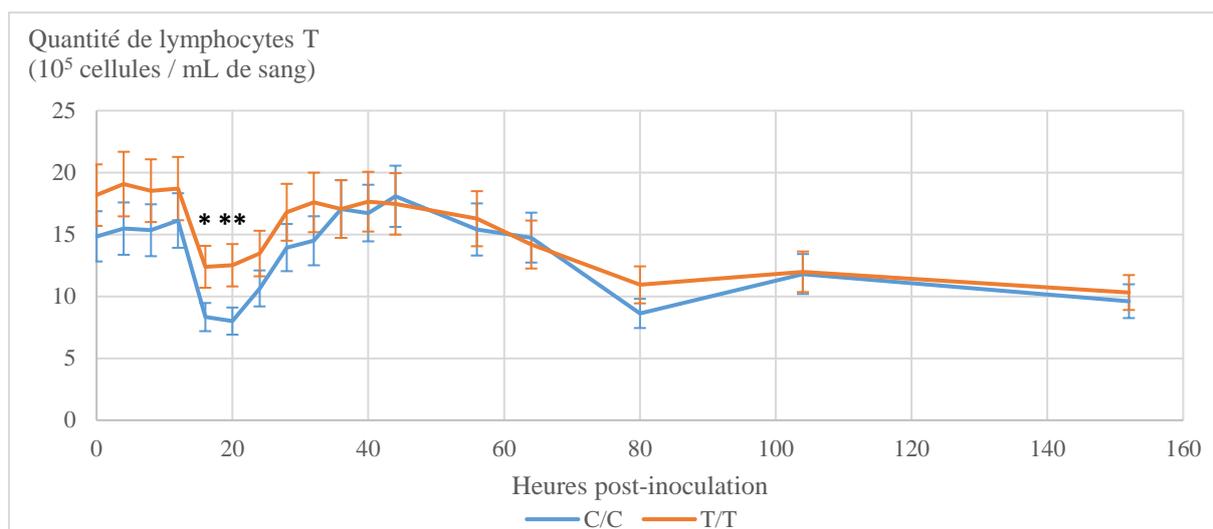


Figure 17 Évolution des lsmeans intra-génotype *SOCS2* de la quantité de lymphocytes T dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05 ; * : p-valeur < 0,10)

Les temps T16 et T20 n'étant pas typés pour les différentes catégories de lymphocytes T (CD4, CD8 et les autres), on ne peut pas confirmer exactement quel type cellulaire semble moins recruté vers les mamelles (potentiellement tous). En revanche, on peut constater que les proportions entre deux de ces trois catégories sont différentes entre les mutants et les sauvages (Annexe 11), et ce, quel que soit le temps du protocole (tendance ; p-valeur effet *SOCS2* global = 0,0627). En effet, les lymphocytes T cytotoxiques (CD8) sont plus nombreux chez les T/T à T0 (tendance ; p-valeur = 0,0861), et le ratio entre CD4 et CD8 est assez constant au cours du temps avec deux temps significatifs : T0 (p-valeur = 0,0405), et T56 (p-valeur = 0,0267) (Figure 18). De ce résultat sort l'hypothèse d'un effet du gène *SOCS2* sur la différenciation des lymphocytes T.

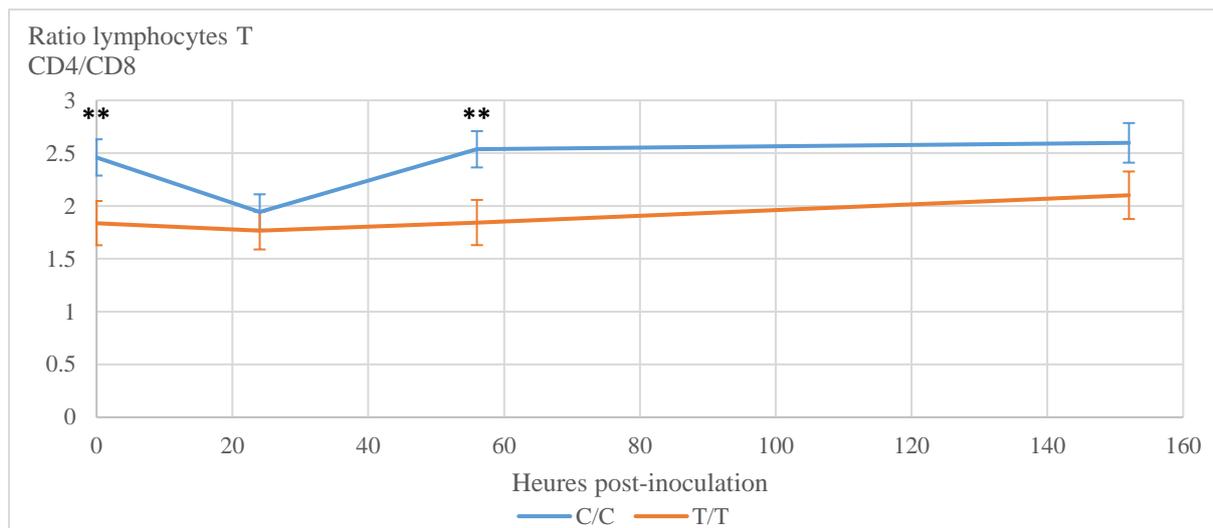


Figure 18 Évolution des lsmeans intra-génotype *SOCS2* du ratio des quantités de lymphocytes T CD4 et CD8 dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05)

Une autre catégorie de cellules immunitaires, les lymphocytes NK (*Natural Killer*), participant à la réponse immunitaire innée de glande mammaire (Sordillo, 2018), semble être impactée par la mutation R96C du gène *SOCS2*, avec une tendance à être moins nombreuse chez les T/T à T152 (p-valeur = 0,0804) (Figure 19). Le rôle de la protéine SOCS-2 sur la différenciation de ces cellules NK avait par ailleurs déjà été rapporté grâce à son inhibition directe sur la protéine JAK-2 (Kim *et al.*, 2017).

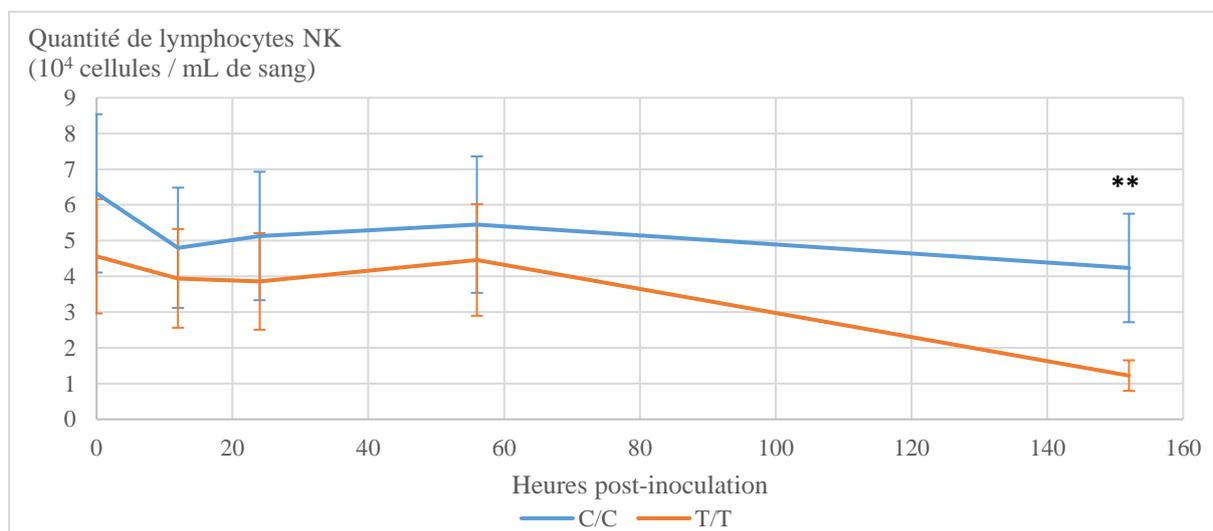


Figure 19 Évolution des lsmeans intra-génotype *SOCS2* du ratio des quantités de lymphocytes NK dans le sang au cours de l'infection expérimentale (** : p-valeur < 0,05)

Parmi les composés solubles (Figure 20) dosés au cours de l'expérimentation, quatre ont retenu notre attention : l'IL-8, pour l'effet global du génotype *SOCS2* sur sa concentration avec plus d'IL-8 chez les T/T (tendance, p-valeur = 0,0932), et pour l'effet significatif à T0 (p-valeur =

0,0324) ; la CCL-2, présente en plus grande quantité chez les C/C à T16 (tendance, p-valeur = 0,0622) ; la CCL-3, présente en moins grande quantité chez les C/C à T8 (tendance, p-valeur = 0,0674) ; et l'haptoglobine, plus concentrée chez les TT au niveau du deuxième pic de recrutement cellulaire à T88 (p-valeur = 0,0037). Ces quatre protéines semblent être intéressantes de par leur rôle primordial dans l'immunité (Palomino *et al.*, 2015). En effet, l'IL-8, la CCL-2 et la CCL-3 sont des chimiokines principales impliquées dans le recrutement respectivement des neutrophiles et des lymphocytes T, des monocytes, et des macrophages. Ceci suggère qu'un dysfonctionnement de la protéine SOCS-2 (mutation à l'étude) provoquerait une dérégulation du trafic cellulaire au moment du recrutement des cellules immunitaires vers le tissu mammaire infecté.

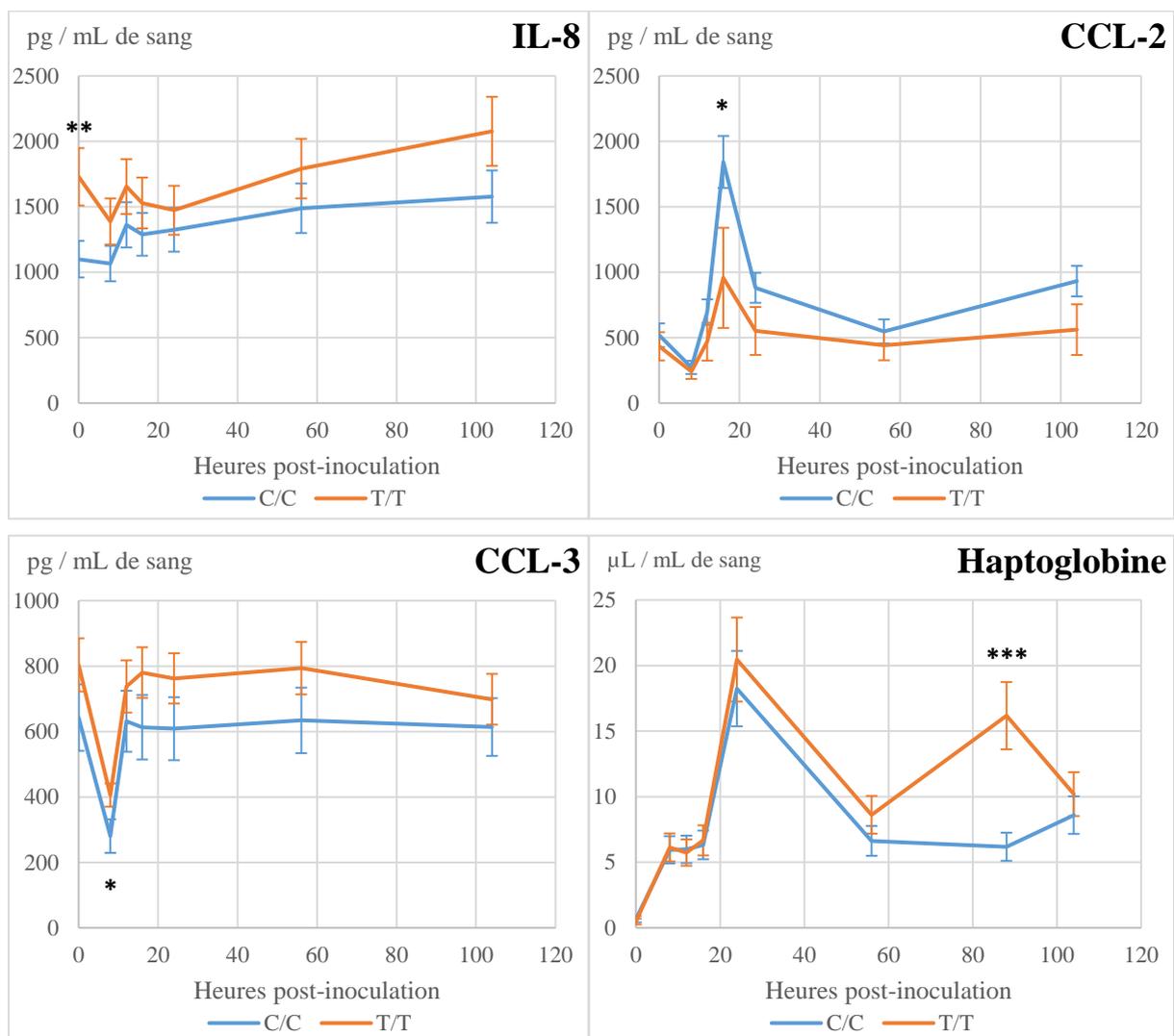


Figure 20 Évolution des lsmeans intra-génotype SOCS2 de la quantité de quelques composés solubles (IL-8, CCL-2, CCL-3, haptoglobine) dans le sang au cours de l'infection expérimentale (*** : p-valeur < 0,01 ; ** : p-valeur < 0,05 ; * : p-valeur < 0,10)

L'haptoglobine quant à elle, est une protéine inflammatoire de la phase aiguë (Eckersall & Bell, 2010). Sa surexpression chez les T/T coïncide avec les différences cliniques observées précédemment (score clinique dégradé chez les T/T) et pourrait s'expliquer comme la conséquence du dysfonctionnement du recrutement des cellules immunitaires (inflammation maintenue) chez les T/T.

III.1.c. Confirmation de l'effet de *SOCS2* sur le gabarit des individus

Sur l'ensemble des seize variables mesurées au moment des autopsies, trois mesures sont significativement différentes entre les brebis C/C et les brebis T/T, celles des os fémur, humérus et radius (Tableau 13). Ce résultat confirme donc ce que nous avons pu voir dans le [Chapitre I](#), sur la régulation de l'hormone de croissance par la protéine SOCS-2 (Metcalf *et al.*, 2000). En effet, les brebis possédant la mutation R96C ont des tailles augmentées de 7,19%, 7,49% et 9,18%, respectivement pour les os fémur, humérus, et radius, par rapport aux brebis sauvages. Le fémur et l'humérus avaient déjà été mesurés précédemment sur des brebis porteuses de la mutation (Rupp *et al.*, 2015) et bien que les différences n'aient pas été signalées comme significatives, les gains des homozygotes mutantes par rapport aux homozygotes sauvages étaient du même ordre de grandeur que celles trouvées dans notre étude (3,9% et 5,8%, respectivement pour le fémur et l'humérus).

Tableau 13 Statistiques descriptives (moyennes et écarts-types) et résultat des tests de comparaison de moyennes sur les variables générées au cours des autopsies de l'infection expérimentale

Variable	Moyenne ± écart-type				P-valeur comparaison de moyennes
	C/C		T/T		
Taille abcès (mm)	1,57 ± 2,07	2,29 ± 1,80	0,539		
Poids vif (kg)	58,79 ± 4,79	62,07 ± 9,06	0,413		
Poids carcasse (kg)	34,57 ± 3,48	35,57 ± 3,83	0,619		
Poids cœur (g)	271,71 ± 36,84	307,86 ± 31,37	0,072		
Poids foie (g)	1086,57 ± 155,67	1077,57 ± 117,22	0,905		
Poids poumons (g)	885,57 ± 264,82	853,86 ± 101,81	0,775		
Poids rate (g)	176,71 ± 56,91	190,57 ± 52,09	0,643		
Poids reins (g)	215,86 ± 50,73	232,00 ± 18,44	0,453		
Poids mamelle inoc. (g)	903,57 ± 338,45	1128,57 ± 508,29	0,318		
Poids mamelle non inoc. (g)	530,57 ± 160,60	525,00 ± 116,94	0,942		
Poids nœud lymphatique inoc. (g)	23,32 ± 10,69	24,88 ± 8,67	0,770		
Poids nœud lymphatique non inoc. (g)	10,08 ± 5,53	8,83 ± 5,42	0,701		
Taille fémur (cm)	19,47 ± 0,70	20,87 ± 0,70	2,76E-03		
Taille humérus (cm)	14,69 ± 0,60	15,79 ± 0,73	9,44E-03		
Taille radius (cm)	16,80 ± 0,57	18,34 ± 0,57	2,84E-04		
Taille tibia (cm)	22,99 ± 1,40	24,21 ± 1,30	0,115		

Abréviation : inoc. : inoculée, mm : millimètre, kg : kilogramme, g : gramme, cm ; centimètre. Les p-valeurs indiquées en gras sont significatives et montrent un effet du génotype de la mutation R96C du gène SOCS2 ovin sur le phénotype.

III.2. Analyses de l'expression génique des cellules sanguines par RNA-seq

En moyenne sur l'ensemble des échantillons, 95,0% des séquences ont été alignées. Sur les 28 échantillons conservés pour la constitution du modèle étendu, 59,6% des lectures alignées en moyenne ont été assignées à un transcrite du modèle de référence. Ce taux d'assignation des lectures alignées passe à 71,4% sur le modèle étendu. Cette valeur peut être considérée comme faible mais est cohérente avec la récente publication du modèle de référence initialement choisi.

Le modèle étendu décrit 140 561 transcrits et 42 461 gènes. Le filtrage réalisé avec *EdgeR* sans prise en compte du génotype *SOCS2* permet de conserver 18 442 gènes comptabilisant au moins 10,7 lectures dans chaque groupe (intra-temps), tandis que celui réalisé avec prise en compte du génotype *SOCS-2* (intra-temps et intra-génotype) permet de conserver 19 832 gènes.

III.2.a. Profilage de la réponse des cellules sanguines à l'infection par *S. aureus*

La réponse immunitaire mobilisée par l'infection expérimentale peut être caractérisée en comparant les gènes différentiellement exprimés entre les différentes dates étudiées dans ce protocole. Les nombres de gènes différentiels entre les temps de l'infection expérimentale avec les deux modèles décrits précédemment (sans considération du génotype *SOCS2*, et intra-génotype) sont présentés dans le Tableau 14.

Tableau 14 Nombre de gènes différentiels entre les temps de l'infection expérimentale (analyse *EdgeR*, *p*-valeur ajustée : 0,05)

<i>Temps</i>	8	12	16	24	56	104
Tout génotypes (Intra-génotype C/C - Intra-génotype T/T)						
0	1 (0 - 0)	0 (0 - 0)	6 278 (4 150 - 2 861)	5 772 (3 252 - 2 324)	501 (4 - 781)	5 120 (2 926 - 1 860)
8		1 (0 - 0)	6 209 (4 041 - 2 815)	5 191 (2 927 - 1 966)	350 (2 - 424)	4 633 (2 689 - 1 587)
12			5 948 (4 050 - 1 779)	6 034 (4 349 - 1 481)	893 (4 - 161)	5 048 (3 467 - 638)
16				2 728 (1 801 - 203)	6 040 (5 057 - 1 987)	3 835 (1 519 - 1 442)
24					3 713 (2 758 - 652)	1 298 (304 - 428)
56						1 967 (2 663 - 0)

Les résultats indiqués en rouge sont ceux pour lesquels une analyse fonctionnelle a été réalisée.

La transcription de gènes est modifiée à partir de T16 avec près de 6 000 gènes différentiels à T16 et T24 (*vs* T0), alors qu'il n'y a pas de gène différentiel entre T0 et T8, T12, ni entre T8 et T12, quelle que soit la méthode d'analyse. Le nombre de gènes différentiels (*vs* T0) diminue ensuite à T56 puis augmente de nouveau à T104 au même ordre de grandeur que pour T16 et T24. Cette évolution de gènes différentiels laisse envisager une visualisation des différents

mécanismes de transcription de molécules de l'immunité mis en place au cours de la réaction infectieuse. Le nombre de gènes différentiels entre T0 et T56 est bien inférieur à ceux entre T0 et tous les autres temps postérieurs à T12 (501 par rapport à plus de 5 000). Cette diminution du nombre de gènes différentiels coïncide avec les diminutions du pic de recrutement de cellules dans le lait (Figure 15) et du statut inflammatoire (cf. haptoglobine Figure 20) observées précédemment.

Globalement, on observe le même profil pour les deux génotypes avec de nombreux gènes différentiels au niveau des deux pics de recrutement cellulaire, bien que ce nombre soit plus important chez les C/C, probablement en raison d'une différence de variabilité entre les deux génotypes (T/T plus hétérogènes).

La répartition moyenne des gènes différentiels entre les analyses en excluant les contrastes où l'on observe moins de 5 gènes différentiels dans au moins une analyse (T0 vs T8 et T12, T8 vs T12, et T56 vs T0, T8, T12 et T104) est présentée en Figure 21. Ces répartitions sont extraites des diagrammes de Venn dont quelques exemples (les quatre temps pour lesquels on observe plus de un gène différentiel vs T0) sont présentés en Annexe 12.

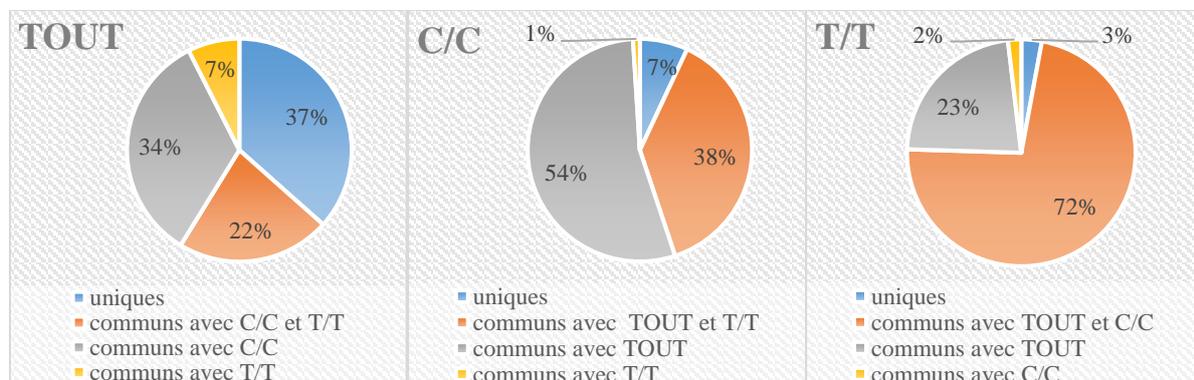


Figure 21 Répartition des gènes différentiels entre les trois analyses (TOUT = tous génotypes, C/C = intra-génotype C/C, et T/T = intra-génotype T/T)

Dans les analyses tous génotypes (TOUT), en moyenne 63% de gènes sont communs (aux deux analyses intra-génotype ou uniquement à une seule). Au vu des faibles pourcentages de gènes différentiels uniques aux analyses intra-génotypes, des analyses fonctionnelles ont été réalisées à partir des résultats des analyses sans considération des génotypes *SOCS2* (contrastes indiqués en rouge dans le Tableau 14). Quatre analyses fonctionnelles ont été réalisées afin de visualiser :

- la réponse immunitaire précoce : contraste T0 vs T16
- la réponse immunitaire au niveau du premier pic d'inflammation : contraste T0 vs T24

- la réponse post-recrutement de cellules immunitaires : contraste T0 vs T56
- la réponse immunitaire tardive (deuxième pic de recrutement de cellules) : contraste T0 vs T104

Les principales voies canoniques (top 10 de chaque analyse) parmi les gènes à expression différentielle de ces quatre contrastes sont présentées en Annexe 13. Tout d'abord, entre T0 et T16, les gènes différentiellement exprimés sont très majoritairement liés à l'activation des voies de la réponse immunitaire adaptative (lymphocytes B, T). Entre T0 et T24, on peut visualiser l'activation d'autres cellules immunitaires en plus des lymphocytes (monocytes, macrophages, cellules NK). Des gènes liés à l'apoptose due à des dégâts sur l'ADN sont également touchés (l'intégrité du squelette est touchée avec la voie de RANK liée au remodelage osseux). A T56, la réponse immunitaire est en place avec de nombreuses voies de réparation de l'ADN, de mitose, et de cycle cellulaire actrices, notamment avec GADD45 qui est induit par la réponse proliférative des lymphocytes T et qui participe aux mécanismes de réparation de l'ADN. Entre T0 et T104, les gènes différentiellement exprimés participent à nouveau au recrutement de cellules immunitaires (lymphocytes). D'autres voies sont impactées, telles que celle jouant sur les interactions cellules-cellules (potentiellement signaux de différenciation).

III.2.b. Une régulation de la réponse immunitaire différente entre les sauvages et les mutants

Le nombre de gènes conservés après filtrage pour cette étude est le même que celui de la seconde analyse du paragraphe précédent (même regroupement), c'est-à-dire 19 832.

En considérant les sept temps du protocole disponibles pour cette analyse, aucun gène différentiellement exprimé n'a été trouvé entre les C/C et les T/T. Ce résultat n'est pas surprenant car à l'échelle du protocole, de nombreux mécanismes biologiques de défense immunitaire se mettent en place, et la protéine SOCS-2 peut potentiellement n'intervenir que temporairement dans un ou plusieurs de ces mécanismes.

D'après les analyses du paragraphe précédent (Tableau 14), on observe moins de gènes différentiels entre les temps chez les T/T, principalement au niveau des deux pics de recrutement de cellules vers la mamelle. En effet, en excluant les contrastes où l'on observe moins de 5 gènes différentiels dans au moins une analyse (T0 vs T8 et T12, T8 vs T12, et T56 vs. T0, T8, T12 et T104), on obtient en moyenne -42% de gènes différentiels chez les T/T par rapport aux C/C. Ceci pourrait potentiellement s'expliquer par une différence de variabilité entre les deux groupes (T/T plus hétérogènes), ou encore par les différences de populations de

cellules sanguines observées précédemment entre C/C et T/T (lymphocytes principalement, cf. Figure 18). La diminution du nombre de gènes différentiels entre T0 et T56 concorde temporellement avec la différence clinique observée entre C/C et T/T (température et score clinique mammaire, respectivement cf. Figure 13 et Figure 14). Par ailleurs, la différence de gènes entre T0 et T56 est plus marquée chez les T/T (781 gènes différentiels) que chez les C/C (4 gènes différentiels), ce qui est pourtant l'inverse de tous les autres temps. On en déduit une potentielle différence de régulation du système immunitaire à un temps particulier (T56).

En testant les contrastes T/T vs C/C à chaque temps, nous avons obtenu 0, 1, 1, 1, 0, 177, et 7 gènes différentiellement exprimés respectivement pour les temps T0, T8, T12, T16, T24, T56, et T104. Ces gènes sont présentés en Annexe 14 avec leur localisation sur le génome, leurs p-valeurs corrigées et leur niveau d'expression entre les T/T et les C/C (fold-change).

Le gène différentiel à T8, sous exprimé chez les T/T (log fold-change = -4,22) est un locus non annoté sur le génome ovin : *LOC114108696*. Le site NCBI l'associe à une protéine ressemblant à l'*ERVK-6* (*Endogenous retrovirus group K member 6* : Rétrovirus endogène membre du groupe K 6) impliquée dans la reconnaissance d'un virus avec son hôte au cours de la réaction immunitaire précoce post-infection. Le gène différentiel à T12 est un gène non annoté provenant du modèle étendu : *XLOC_000401*. Il est localisé sur un contig, c'est-à-dire une séquence génomique continue et ordonnée non alignée sur le génome de référence. Il n'est donc pas possible à l'heure actuelle de connaître la protéine codée par ce gène. Le gène différentiel à T16 est également un gène non annoté provenant du modèle étendu : *XLOC_010460*. Cependant, l'expression de ce gène serait localisée sur le chromosome 13 entre 49 754 573 et 49 771 981 paires de bases, région dans laquelle aucun gène n'est en effet annoté sur le génome de référence.

Parmi les 7 gènes différentiellement exprimés à T104, seulement 3 sont annotés (*GPAT2*, *FZD6* et *APIM2*), deux sont non annotés mais décrits par le modèle de référence (*LOC114110327* et *LOC114118421*) et les deux derniers sont des loci du modèle étendu (*XLOC_000627* et *XLOC_000330*), localisés sur des contigs.

De façon notable, à T56, 177 gènes sont différentiellement exprimés entre les C/C et les T/T : 158 surexprimés et 19 sous-exprimés chez les T/T. Parmi eux, 14 sont des loci englobant plusieurs gènes. Parmi les 163 restants, 82 sont annotés par le modèle de référence, les autres (81) sont des *loci* non annotés soit décrits par le modèle de référence (40 commençant par « *LOC* »), soit décrits par le modèle étendu (41 commençant par « *XLOC* »). Dans Ingenuity,

80 des 82 gènes annotés sont reconnus par Ingenuity (annotation chez l'homme) : *BST-2A* (le « a » non connu chez l'homme) et *TRNAC-CAC* (ARN transférase) sont manquants. Après annotation manuelle (site NCBI), Ingenuity reconnaît 113 gènes sur 177.

Les voies les plus représentées d'après l'analyse Ingenuity (Tableau 15) sont les voies de signalisation des interférons (suractivée chez les T/T) et de STAT-3. Les IFN- α et - β ont déjà été rapportés comme étant régulés par SOCS-2 au cours d'une infection chez des souris par le virus BoHV-5 (Bovine herpesvirus 5 : herpès-virus bovin 5) (Aparecida Silva Barbosa *et al.*, 2016). Les souris déficientes en SOCS-2 réagissaient avec une inflammation exacerbée et présentaient une expression accrue de l'ARN messager de SOCS-1 par rapport aux souris de type sauvage. L'IFN- γ est également connu pour être régulé par la protéine SOCS-1 (Starr *et al.*, 2009). Nous pourrions donc émettre une hypothèse de régulation de SOCS-1 par SOCS-2 dans notre expérimentation. L'action directe ou indirecte de SOCS-2 ne peut cependant pas être proposée avec ces données.

Tableau 15 Principales voies canoniques (ratio < 0,05 et q-valeur < 0,05) parmi les gènes à expression différentielle entre les C/C et les T/T à T56 de l'infection expérimentale d'après le logiciel Ingenuity Pathway Analysis

<i>Voie canonique</i>	<i>Ratio</i>	<i>q-valeur</i>	<i>Sur/Sous</i>	<i>Gènes à expression différentielle régulé</i>
<i>Signalisation des interférons</i>	0,194	1,3E-07	6/0	<i>IFI35, IFI6, IFIT3, IFITM1, IFITM3, MX1, OAS1</i>
<i>STAT3</i>	0,067	3,4E-06	9/0	<i>CSF2RB, IL18R1, IL18RAP, IL1R2, IL22RA2, IL2RA, IL3RA, IL4R, RALB</i>

Le ratio est le rapport entre le nombre de gènes différentiels sur le nombre de gènes de la voie. La q-valeur montre l'importance de l'enrichissement d'une fonction au sein des gènes différentiels, ajustée par le FDR de Benjamini et Hochberg.

Les régulateurs mis en avant dans cette analyse sont pour la grande majorité des facteurs de transcription suractivés chez les mutants (top 10 : IRF7, STAT3, IRF3, IRF1, STAT1, IRF5, STAT2, CNOT7). Le manque de la protéine SOCS-2 fonctionnelle, une protéine inhibitrice d'une voie de signalisation, joue sur la transcription de gènes à un moment où le recrutement cellulaire semble être terminé, il pourrait donc bien s'agir du rétrocontrôle de ce mécanisme de recrutement qui est modifiée par la mutation.

III.3. Intégration de l'ensemble des données

III.3.a. A l'échelle du protocole, des différences fortes entre temps et plus faibles entre génotypes

Parmi l'ensemble des PCA réalisées sur le premier jeu de données (données disponibles pour tous les temps du transcriptome), aucune ne permet de différencier les individus C/C des individus T/T, quel que soit le bloc de variables (cliniques, lait-sang, et comptage). En revanche, on peut distinguer quelques clusters qui différencient les temps entre eux, et notamment avec le bloc de données de comptages de gènes qui permet une nette séparation des échantillons de T16 et T24 (Figure 22). Par ailleurs, l'ajout de l'effet de répétition permet une explication de la variabilité du jeu de données plus grande sur les deux premières composantes (on passe de 17,4% et 9,2% sans effet de répétition à 21,8% et 11,1% de variance expliquée avec effet de répétition, respectivement pour les deux premières composantes), et une meilleure visualisation des clusters plus homogènes. De plus, les 10 gènes conservés avec l'option « *sparse* » sur le premier axe qui sépare les temps T16 et T24 des autres (non montré, 20,4% de variance expliquée) sont majoritairement des gènes liés à des transmissions de signal cellulaire (*SLC2A3*, *FAM129A*, *RALB*, *PICALM/XLOC_023279*, *GK*, *ACVR1B*, *CACNB2*, *FGR*, *XLOC_041088*, *LRG1*). D'après ces PCA, la source principale de variabilité du jeu de données à l'échelle du protocole provient donc des différences entre les temps de contrôle, c'est-à-dire de la réponse immunitaire.

Ceci est confirmé par les PLS-DA avec comme facteur discriminant le « temps de contrôle » (Figure 23, A). En effet, les pourcentages d'explication de variance du jeu de données de comptages sont très proches de ceux des PCA (21,5% et 9,4% pour les deux premières composantes). De plus, les 10 gènes conservés avec l'option « *sparse* » sur le premier axe qui sépare les temps T16, T24 et T104 des autres (non montré, 17,7% de variance expliquée) sont également majoritairement des gènes liés à des transmissions de signal cellulaire, mais plus spécifique de la réponse immunitaire (*LOC114109421/MYD88*, *TLE3*, *PIMI*, *E2F2*, *SCARB1*, *UBALD2*, *MTF1*, *STAT3*, *SBNO2*, *ARID5A*).

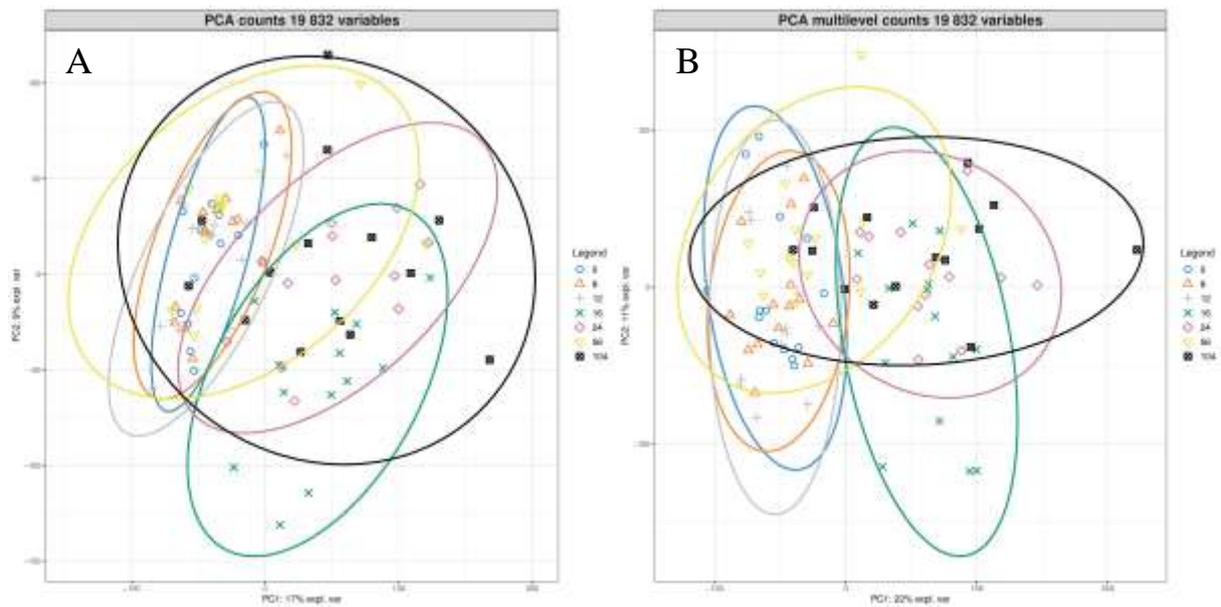


Figure 22 PCA sur le bloc des données de comptages de gènes à tous les temps du transcriptome de l'infection expérimentale, avec (A) et sans (B) prise en compte de l'effet animal (option multilevel)

Chaque temps de contrôle est matérialisé par une couleur et un symbole différents, comme indiqué dans la légende. Chaque point symbolisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole.

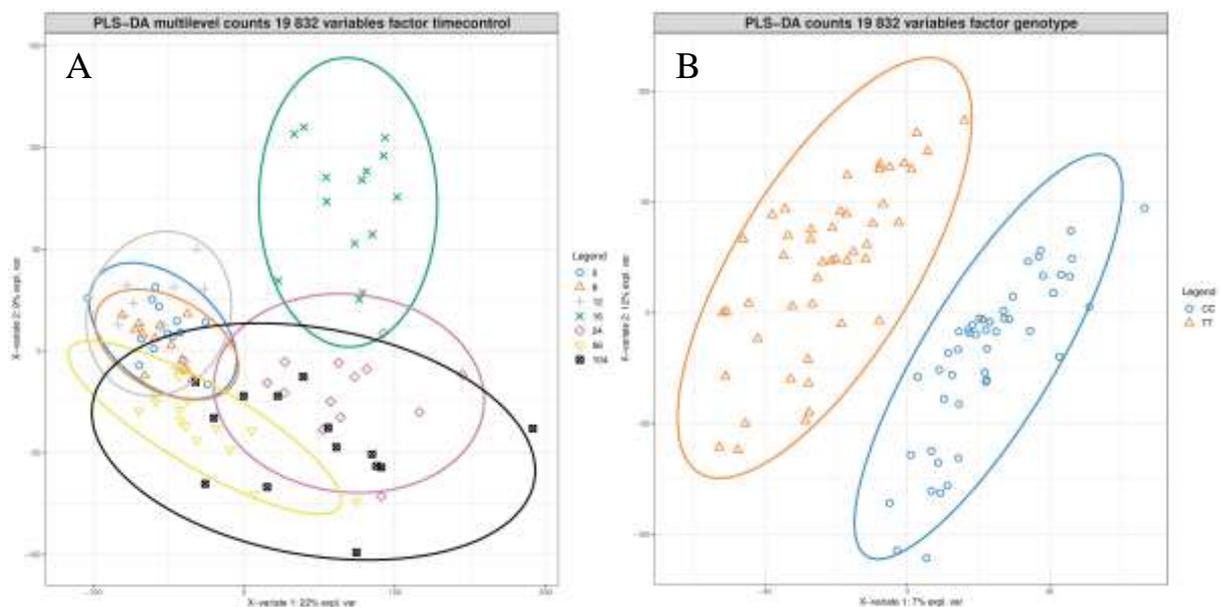


Figure 23 PLS-DA avec comme facteurs discriminants le temps de contrôle (A) et le génotype *SOCS2* (B) sur le bloc des données de comptages de gènes à tous les temps du transcriptome de l'infection expérimentale

Chaque temps de contrôle (A) ou chaque génotype *SOCS2* (B) est matérialisé par une couleur et un symbole différents, comme indiqué dans la légende. Chaque point symbolisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole.

Malgré cette forte variabilité expliquée par la réponse immunitaire, la PLS-DA avec comme facteur discriminant le « génotype *SOCS2* » permet de faire ressortir une différence entre les C/C et les T/T (Figure 23, B). En effet, l'axe 1 permet une nette distinction entre les deux clusters pour une explication de la variance de 7,0%. Sur les 10 gènes conservés sur ce premier

axe avec l'option « *sparse* » (non montré, 2,5% de variance expliquée), 8 d'entre eux font partie des gènes différentiellement exprimés entre C/C et T/T dans les analyses décrites précédemment (un à T12, six à T56, et un à T104) : *LOC114112166*, *XLOC_000627*, *XLOC_000401*, *XLOC_035910*, *LOC114112479/XLOC_000758*, *XLOC_000330*, *XLOC_035892*, *LOC101103588*, *LOC114112552/LOC114112554*, *XLOC_000906*. Malheureusement, aucun de ces gènes n'est annoté.

Les multi-bloc PLS-DA montrent le même résultat avec de bonnes corrélations entre le facteur discriminant « temps de contrôle » (options « *sparse* » et « *multilevel* ») et les trois blocs de variables quantitatives (0,81, 0,62, et 0,66 respectivement pour les blocs de comptages, cliniques et lait-sang). Les deux premiers axes expliquent 20,1% et 7,2%, 45,5% et 19,9%, et 24,4% & 11,9%, respectivement pour les blocs de comptages, cliniques et lait-sang, et permettent pour les trois blocs la distinction nette des différents temps de contrôle (Annexe 15).

Les corrélations avec le facteur discriminant « génotype *SOCS2* » (option « *sparse* » uniquement) sont plus faibles qu'avec le facteur discriminant « temps de contrôle » : 0,49, 0,15 et 0,34 respectivement pour les blocs de comptages, cliniques et lait-sang. Les deux premiers axes expliquent 15,4% et 3,0%, 42,2% et 16,0%, et 15,0% et 15,4%, respectivement pour les blocs de comptages, cliniques et lait-sang (Annexe 16). La distinction entre les deux génotypes est bien nette pour le bloc des données de comptage et de lait-sang, et peu visible pour les données cliniques.

III.3.b. Des gènes associés à la différence de régulation liée à *SOCS2*

À T56, seules les données de comptages permettent une très nette distinction des génotypes *SOCS2* sur les PLS-DA, et ce malgré la plus grande hétérogénéité des T/T (Annexe 17). Les deux premiers axes de la PLS-DA sur les données de comptage expliquent respectivement 13,3% et 7,9% de la variance. Bien que la séparation des génotypes se visualise sur l'axe 1, parmi les 10 gènes conservés (*LOC101117206*, *KLRG2/LOC105609822*, *XLOC_033871*, *CCNF*, *C4BPB*, *PTPRF*, *XLOC_031753*, *SEC23IP*, *ZBTB39*, *XLOC_008939*) maximisant la covariance entre le facteur discriminant et le bloc des variables de comptages (option « *sparse* »), seulement deux gènes sont différentiellement exprimés entre les C/C et les T/T (*XLOC_031753* et *XLOC_008939*) d'après les analyses précédentes (cf. III.2.b).

Les PLS-DA sur les données cliniques et sur les données du sang permettent également une légère distinction des génotypes sur les axes 1 qui expliquent respectivement 55,9% et 12,9% de la variance pour les données cliniques et celles du sang. Les variables qui semblent expliquer

le plus cette différence sont la température du rumen imputée, et le score clinique de la mamelle inoculée pour les données cliniques et le rapport entre lymphocytes T CD4 et CD8, l'haptoglobine, et les lymphocytes T CD8 pour les données du sang, validant ainsi les résultats des modèles mixtes observés précédemment (cf. III.1.b).

Avec la *sparse* multi-bloc PLS-DA, on obtient des corrélations raisonnables entre le facteur discriminant « génotype *SOCS2* » et les quatre blocs de variables quantitatives : 0,57, 0,45, 0,35, et 0,37 respectivement pour les blocs de comptages, cliniques, lait, et sang. Les deux premiers axes expliquent 19,9% et 14,0%, 59,9% et 6,1%, 45,3% et 25,0%, et 25,3% et 21,4%, respectivement pour les blocs de comptages, cliniques, lait, et sang.

Les génotypes se distinguent sur les axes 1 (Annexe 18) qui sont principalement expliqués par la température du rumen imputée, le score clinique général, la fréquence cardiaque et le score clinique de l'hémi-mamelle inoculée pour le bloc des données cliniques, par le nombre de neutrophiles morts, et le nombre de lymphocytes pour le bloc des données du lait, et par la quantité d'IL-6, le nombre de monocytes inflammatoires, et la quantité d'IL-1 β pour le bloc des données du sang (Annexe 19). Ces résultats sont cohérents avec ceux obtenus précédemment par l'approche des modèles mixtes, c'est-à-dire principalement une inflammation et une dégradation de l'état clinique accrues des T/T à T56. De plus, sur les 10 gènes conservés sur le premier axe, 6 d'entre eux font partie des gènes différentiellement exprimés entre C/C et T/T dans les analyses décrites précédemment à T56 : *DPH5*, *XLOC_020177*, *IL18RAP*, *LOC114113097/XLOC_003973*, *GYG1*, *XLOC_039743*.

Cette approche met également en lumière des variables qui n'étaient pas significatives précédemment, telles que l'IL-6 par exemple où l'on observait pourtant une différence de moyenne conséquente entre C/C et T/T (plus chez les T/T) à T56 (Annexe 10). L'IL-6 est une cytokine impliquée dans l'inflammation activant la protéine STAT-3, et est régulée par la protéine SOCS-3 (Schaper & Rose-John, 2015), ce qui concorde avec les résultats de l'analyse Ingenuity obtenus précédemment. Sa concentration est par ailleurs très corrélée (corrélation supérieure à 0,95) au gène *IL18RAP* que l'on retrouvait parmi les gènes différentiels appartenant à la voie STAT-3 identifiée (Figure 24).

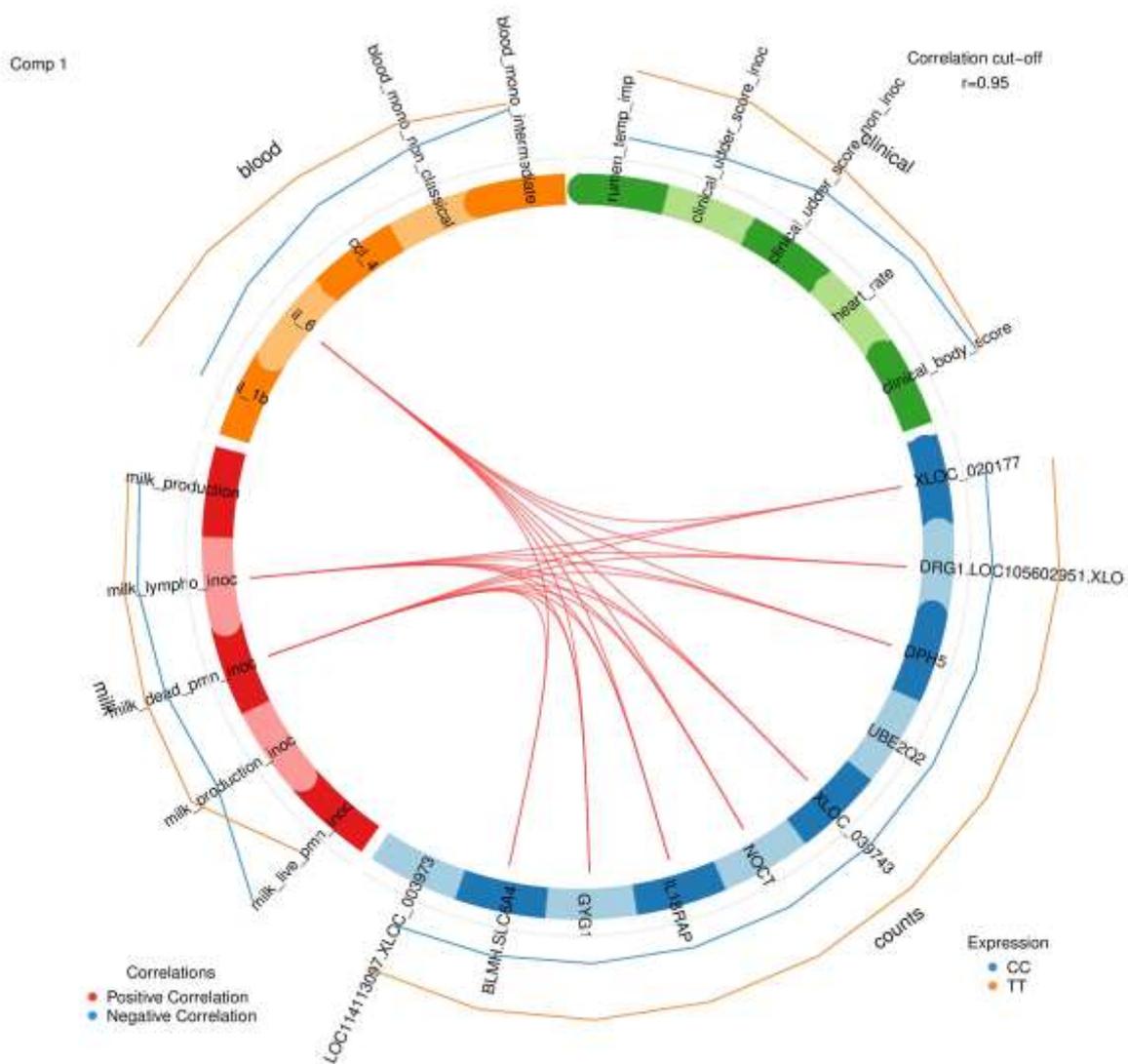


Figure 24 Cercle des corrélations entre les variables expliquant le plus l'axe 1 de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » pour chaque bloc : comptages de gènes (en bleu), données cliniques (en vert), données du lait (en rouge) et données du sang (en orange) à T56 de l'infection expérimentale

IV. Conclusions du Chapitre IV.

Dans ce chapitre, nous cherchions à déterminer plus précisément le rôle de la protéine SOCS-2 dans les mécanismes biologiques sous-jacents de l'immunité grâce à l'étude de la réponse de brebis porteuses de la mutation et non porteuses de la mutation R96C dans le cadre d'une infection intra-mammaire par un staphylocoque (*S. aureus*). De nombreuses données phénotypiques, cliniques et transcriptomiques ont été collectées et générées au cours de cette infection jusqu'à 152 heures post-inoculation.

L'approche *via* les modèles mixtes sur les variables phénotypiques et cliniques a montré une réaction immunitaire chez l'ensemble des brebis, avec deux pics d'inflammation et de recrutement de cellules immunitaires passant du sang vers la mamelle : un aux alentours de 20 heures post-inoculation, et un autre aux alentours de 88 heures post-inoculation. La réaction clinique était accrue chez les T/T avec notamment une différence significative entre les deux pics (T56) pour les variables de température et de score clinique mammaire côté inoculé. Les modèles mixtes ont également mis en évidence une différence de recrutement de cellules immunitaires et plus particulièrement de lymphocytes T au niveau des pics.

L'analyse différentielle des gènes exprimés à sept temps clés du protocole a confirmé pour l'ensemble des individus une modification du transcriptome à partir de T16 (début de pic inflammatoire) en lien avec l'activation du système immunitaire. Les gènes exprimés différemment sont des gènes participant à l'activation des différentes cellules immunitaires dans un premier temps, puis dans un second temps, à partir de la fin de la première vague de recrutement cellulaire, les exprimés sont des gènes des voies de réparation de l'ADN et d'apoptose cellulaire. On observe des gènes différemment exprimés entre les C/C et les T/T uniquement à T56, au moment où les cellules immunitaires sont actives et où l'on peut faire l'hypothèse que le phénomène est régulé, concordant ainsi avec le rôle de la protéine SOCS-2 dans la voie JAK/STAT. Par ailleurs, une analyse fonctionnelle sur ces gènes différentiels a montré que les voies suractivées chez les mutants sont bien des voies de transmission de signal et de régulation (interférons, STAT-3).

Les analyses d'intégration ont conforté ces résultats avec des variables les plus explicatives de la variabilité des jeux de données qui sont en accord avec les résultats des modèles mixtes, ainsi que de l'analyse différentielle. Des analyses complémentaires seraient nécessaires afin d'investiguer plus minutieusement les voies modifiées par la mutation, et plus particulièrement les différents facteurs de transcription activés en l'absence d'une protéine SOCS-2 fonctionnelle.

Chapitre V. Discussion générale et perspectives

Les résultats de cette thèse ayant déjà fait l'objet de discussions dans chacun des chapitres de ce manuscrit, nous allons faire ici un bref rappel des principales conclusions et nous les resituerons dans le contexte actuel de la sélection chez les ovins en France mais également dans un contexte de recherche plus large. Des perspectives aux travaux seront également présentées.

I. Un effet de *SOCS2* sur la reproduction ?

Dans le [Chapitre III](#), nous avons montré que les brebis homozygotes pour la mutation *SOCS2* avaient un risque d'échec à l'IA 1,3 fois supérieur aux brebis homozygotes sauvages. En revanche, le nombre de petits par portée n'était pas impacté chez les brebis fécondées. Ces résultats sont confortés par une analyse populationnelle récente qui a montré un déficit d'homozygotes dans la région du gène *SOCS2* pour les béliers génotypés pour la sélection génomique (Maxime Ben Braiek, communication personnelle). Par ailleurs, nous avons constaté la difficulté à procréer des individus T/T au domaine de La Fage, mais les fréquences des différents génotypes n'étaient pas significativement différentes des effectifs mendéliens attendus. Les données accumulées au cours de la thèse ne nous permettaient pas toutefois de déterminer à quel stade physiologique (production et fonctionnalité des gamètes chez le mâle ; ovulation, fécondation, implantation et développement embryonnaires, etc. chez la femelle) la protéine SOCS-2 pouvait intervenir dans la régulation de ce phénomène complexe qu'est la reproduction.

Chez la brebis, l'activité sexuelle est saisonnière (jours courts, automne-hiver) et se caractérise par des cycles œstraux d'une durée moyenne de 17 jours. Ces cycles se décomposent en deux phases : folliculaire (3 à 4 jours) et lutéale (13 à 14 jours). La phase folliculaire est une phase de croissance des follicules, structures sphériques présentes dans l'ovaire contenant l'ovocyte, le gamète femelle potentiellement fécondable. Cette phase se caractérise par une production croissante d'œstradiol et se termine par les chaleurs (ou œstrus) et le phénomène d'ovulation, *i.e.* la libération d'un ou plusieurs ovocytes hors de l'ovaire dans les voies génitales. La phase lutéale est une phase où un follicule ayant libéré son ovocyte devient une structure lutéale (ou corps jaune) et sécrète fortement une autre hormone, la progestérone. Si la brebis n'est pas fécondée, cette phase lutéale est interrompue par régression de la (des) structure(s) lutéale(s) et laisse place à un nouveau cycle (une nouvelle phase folliculaire).

Il serait envisageable d'apporter plus d'information quant au rôle que joue SOCS-2 dans ce phénomène de reproduction grâce au dispositif expérimental actuellement en place au sein de l'unité expérimentale INRA La Fage. En effet, à La Fage, des individus porteurs et non porteurs de la mutation *SOCS2* sont produits et suivis pour différents protocoles de recherche. Les brebis sont mises à la reproduction chaque année au cours de l'été pour une mise-bas prévue en novembre-décembre. Actuellement, le suivi de la reproduction se fait exclusivement par un diagnostic de gestation par échographie transabdominale deux à trois fois au cours de la gestation, ce qui permet de détecter les brebis non fécondées ou ayant avorté, et ainsi de décider de la conduite à adopter (écartement de la brebis, remise à la reproduction).

Une première hypothèse concernant le rôle de SOCS-2 dans la fertilité femelle serait que le système endocrinien induisant l'ovulation (axe somatotrope) est dérégulé par la mutation. Il a été montré en bovin que la voie JAK/STAT et notamment la protéine SOCS-2 semblait être impliquée dans la régulation de ce phénomène d'ovulation. En effet, d'une part, la voie JAK/STAT, et plus particulièrement la protéine JAK-3, voit son expression diminuer dans la phase terminale du développement folliculaire en particulier lors du pic pré-ovulatoire d'hormone lutéinisante (LH) (Ndiaye *et al.*, 2016). D'autre part, l'expression de *SOCS2* atténuerait la voie de transduction du signal de l'hormone de croissance hypophysaire *via* son récepteur, au moment de l'augmentation des taux d'œstradiol juste avant l'ovulation des génisses (Mense *et al.*, 2015). Par ailleurs, le récent atlas ovin d'expression génique montre que *SOCS2* est bien exprimé dans l'ovaire adulte (Clark *et al.*, 2017). L'étape d'ovulation pourrait être suivie plus finement à La Fage par des échographies transrectales ou par coelioscopie grâce au dénombrement précis des corps jaunes. Cette mesure permettrait de déterminer si la protéine SOCS-2 a une influence sur le taux d'ovulation.

Une seconde hypothèse concernant le rôle de SOCS-2 serait au niveau de l'endomètre et du placenta après ovulation : effet néfaste de la mutation sur la reconnaissance maternelle, influences du génotype maternel et/ou fœtal. En bovin, des gènes (dont les *SOCS*) ont été identifiés comme marqueurs du phénomène de reconnaissance maternelle et d'implantation (Carvalho *et al.*, 2014). En outre, le placenta est le tissu dans lequel *SOCS2* est le plus exprimé d'après le récent atlas ovin d'expression génique (Clark *et al.*, 2017). Autour du moment de l'ovulation, des prélèvements de sang pourraient être effectués régulièrement à La Fage afin de suivre précisément l'évolution des différentes hormones tels que l'œstradiol ou encore la progestérone. Sur ces prélèvements pourraient également être effectuées des analyses transcriptomiques qui devraient mettre en lumière les gènes importants dans ce phénomène. De manière moins invasive, des échographies transabdominales plus rapprochées à partir du 15^{ème}

jour de gestation (après l'implantation) devraient permettre de déterminer si la reconnaissance maternelle de la gestation et l'implantation embryonnaire a bien été effectuée, et suivre plus finement les cas d'avortement s'il y en a, à intégrer avec les données sanguines pour expliquer les raisons de l'avortement.

L'effet potentiellement néfaste de la mutation *SOCS2* sur la reproduction est important à estimer pour la filière en raison des coûts supplémentaires engendrés par une brebis non fécondée au cours d'une campagne de reproduction, que ce soit dans la filière laitière ou dans la filière bouchère. Dans la population Lacaune au contrôle laitier, le taux de réussite à l'IA est actuellement de 68% (communication personnelle, Rachel Rupp) et l'on peut estimer la fréquence des brebis C/T à environ 20% de la population. Nous avons trouvé dans ce travail de thèse que les brebis C/T avaient un risque d'échec à l'IA 1,06 fois supérieur en comparaison des brebis de type sauvage. Si la mutation était éradiquée dans les élevages, ce taux moyen de réussite à l'IA pourrait donc potentiellement dépasser rapidement les 70% d'après notre étude. Actuellement, la reproduction ne fait pas partie des caractères en sélection en Lacaune mais des réflexions sont en cours pour la prise en compte de phénotypes liés à ce caractère dans les index, voire l'implémentation d'informations de marqueurs génétiques associés (gènes de fertilité vus dans le [Chapitre I.](#), gène *SOCS2*, ou autres QTL avec des méthodes pondérées), pour répondre aux attentes de la filière.

II. Vers une sélection avec prise en compte de la mutation ?

La question de la gestion de cette mutation au sein de la population, favorable pour les caractères de production, et défavorable pour les caractères de santé et de reproduction n'a pas encore été clarifiée par la filière. En effet, depuis 2012, des recommandations ont été émises par les collectifs de recherche en faveur d'une diminution de la fréquence des homozygotes T/T parmi les mâles améliorateurs afin de limiter la propagation de l'allèle muté. Mais l'éradication de cette mutation n'est pas aujourd'hui d'actualité.

Dans le [Chapitre II.](#), nous avons montré que l'allèle muté était de moins en moins présent chez les mâles des stations d'amélioration, et ce, depuis 2005. Ce résultat coïncide avec l'entrée du caractère CCS parmi les caractères en sélection afin d'améliorer la santé mammaire. Ceci est dû au fait que le *locus* de la mutation est en forte association avec ce caractère lié à la santé mammaire comme nous l'avons montré dans cette thèse avec une variance génétique de ce caractère expliquée pour 12% par la région de la mutation. Par ailleurs, les résultats du modèle *Gene Content* testé dans ce travail de thèse, permettant de dissocier la partie génétique due au gène *SOCS2* et la partie génétique due à la composante polygénique, mettent en lumière le fait

que la sélection sur les CCS permet un gain global en faveur d'une résistance aux mammites qui serait plus dû à l'effet polygénique qu'à l'effet de la mutation *SOCS2*. Par conséquent, la sélection sur la résistance aux mammites est actuellement suffisante pour faire diminuer la fréquence de cette mutation défavorable sans pour autant perdre le progrès génétique obtenu sur les autres caractères de production. Il n'est donc probablement pas nécessaire de mettre en place une éradication totale de cette mutation. Les effets et la fréquence de cette mutation pourraient par ailleurs être évalués dans la filière allaitante au sein de laquelle les caractères de croissance (poids, développement musculaire, engraissement) sont de grand intérêt, représentant environ 30% des objectifs de sélection en Lacaune (communication personnelle, Flavie Tortereau), et où les problèmes de mammites sont de plus en plus au cœur des réflexions (O'Brien *et al.*, 2017 ; McLaren *et al.*, 2018). La puce de basse densité (18K) ovine (non encore en place en Lacaune) contient par ailleurs la mutation *SOCS2*, et pourrait être utilisée pour estimer sa fréquence dans les différentes populations (communication personnelle, Flavie Tortereau). D'autant plus que cette mutation semble intéresser les chercheurs à travers le monde pour son effet principal sur la croissance d'après les résultats de l'édition récente d'un génome ovin de race allaitante pour cette mutation (Zhou *et al.*, 2019).

Actuellement en France, les mâles des stations de testage sont tous génotypés avec la puce de basse densité (18K) qui contient le *locus* de la mutation *SOCS2*. Cette puce est imputée vers la puce de moyenne densité utilisée pour la sélection et dont le *locus* de la mutation ne fait pas partie. Les gains en précision des prédictions des évaluations génomiques, apportés par les méthodes pondérées testées dans cette thèse, sont intéressants, notamment pour les caractères les plus associés (CCS, quantité de matière protéique, profondeur de la mamelle). Les méthodes pondérées pourraient donc être appliquées afin d'augmenter les précisions des évaluations et ainsi l'efficacité du schéma de sélection.

III. Les GWAS à partir des modèles de prédiction ?

Les méthodes conventionnelles pour la cartographie du déterminisme génétique des caractères testent, dans un modèle de régression, des marqueurs moléculaires un à un pour leur association au caractère en question. Les phénotypes sont pré-corrigés pour les effets environnementaux, et l'on corrige les résultats des tests pour les associations multiples ultérieurement. Avec les avancées technologiques, le nombre de marqueurs génotypés et testés a considérablement augmenté ces 10 dernières années, de même que le nombre de marqueurs déclarés significatifs (Wood *et al.*, 2014). Il devient alors difficile de conclure quant au SNP réellement causal parmi tous les marqueurs en déséquilibre de liaison avec un QTL avec cette approche.

Les méthodes utilisées dans cette thèse mettent en lumière une autre approche pour la recherche de QTL. En Lacaune lait, à partir des résultats de (Duchemin *et al.*, 2012) qui montraient que les méthodes BLUP étaient équivalentes aux méthodes bayésiennes en terme de précision des prédictions génomiques, les modèles génomiques en routine se sont développés à partir des méthodes BLUP basées sur le pedigree déjà en place. L'approche BLUP en une seule étape (« *single-step* ») a été montrée ultérieurement plus précise grâce à l'ajout de l'information phénotypique d'individus non génotypés sans étape de correction préalable (Baloche *et al.*, 2014). Nous avons donc testé dans cette thèse différentes approches *single-step* BLUP génomiques qui permettaient entre autres d'extraire l'information des poids des marqueurs à partir de la décomposition des valeurs génétiques au cours de plusieurs itérations d'évaluation (Wang *et al.*, 2012 ; Zhang *et al.*, 2016). Ces poids correspondent à l'effet des marqueurs sur le caractère prédit. Cela nous a ainsi permis de valider et confirmer des QTL connus, et d'en découvrir d'autres avec un seuil d'explication de la variance génétique de 1%. Notre étude ne portant pas sur la recherche de QTL, nous avons peu approfondi cet aspect de ces approches, pourtant très intéressant.

Les méthodes de prédiction BLUP peuvent donc aujourd'hui apporter de l'information sur le poids qu'un marqueur a sur la détermination du caractère prédit. L'avantage de ces approches est qu'elle permet d'une part l'estimation conjointe des effets fixes et les effets des marqueurs, et d'autre part, l'ajout d'information phénotypique car tous les individus phénotypés (souvent les filles pour les caractères laitiers) ne sont pas nécessairement génotypés (information génétique du père) grâce aux approches « *single-step* ». Cet ajout d'information rend possible la réalisation de GWAS avec les données collectées en routine, sans génotypage supplémentaire. De plus, ces approches sont plus faciles à implémenter car l'on part directement des données brutes, et plus précises (Wang *et al.*, 2012).

Les approches bayésiennes d'évaluation génomique, telles que Bayes R (Erbe *et al.*, 2012), permettent également d'identifier les polymorphismes qui affectent les caractères (Goddard *et al.*, 2016). Cette méthode serait plus efficace et plus précise que les méthodes BLUP car elle permet d'éliminer les SNPs non informatifs et d'identifier plus précisément les polymorphismes causaux. Cependant, cette approche était réalisée en deux étapes et ne permettait pas l'inclusion de phénotypes d'individus non génotypés. Récemment, des approches bayésiennes en une seule étape ont été développées (Fernando *et al.*, 2014) et commencent à être testées en sélection (Gao *et al.*, 2018). Ces approches sont prometteuses pour la recherche de QTL avec des données de séquence par exemple.

IV. Les polymorphismes de *SOCS2*

La mutation du gène *SOCS2* étudiée dans cette thèse est localisée en position 96 de la protéine mature, située dans le domaine connu de liaison à ses ligands (Kung *et al.*, 2019). Cette localisation est responsable de la non-fonctionnalité de la protéine et de l'effet pléiotrope engendré sur un grand nombre de phénomènes régulateurs. Cette même mutation a été trouvée dans la race ovine française Manech Tête Rousse (Gwenola Tosser-Klopp, communication personnelle). Pour le moment, aucune étude n'a été réalisée pour connaître l'origine de ce partage de variants (origine commune ou introduction accidentelle et ultérieure de Lacaune), mais il serait intéressant de l'investiguer. En effet, s'il s'avère que cette mutation est apparue chez un ancêtre commun de l'espèce ovine, la recherche systématique et la gestion de ce gène majeur dans les autres schémas de sélection pourrait être à prendre en compte.

Par ailleurs, d'autres variants existent dans ce gène chez les ruminants d'après les bases de données du site internet Ensembl. A ce jour, en caprin, 26 variants sont connus, dont 2 sont des mutations ponctuelles provoquant des modifications d'acides aminés en positions 17 et 84 de la protéine mature. En ovin, 28 variants sont documentés, dont 3 sont des mutations ponctuelles (positions 10, 96 et 144, celle de la position 96 étant celle à l'étude). En bovin où les études abondent, Ensembl recense 351 variants dont 40 sont des mutations ponctuelles. Parmi elles, quelques-unes sont proches de la mutation étudiée dans cette thèse, en position 89 ou encore en position 98. Toutes ces mutations sont susceptibles d'avoir des effets similaires à celle étudiée ici et mériteraient d'être investiguées.

V. Effet de la mutation *SOCS2* sur la réponse immunitaire

Identifiée grâce à des études d'association tout génome pour le caractère CCS, la mutation *SOCS2* étant dans un premier temps principalement reliée à l'inflammation mammaire (Rupp *et al.*, 2015). Ces résultats obtenus dans des élevages commerciaux, au sein desquels les animaux sont exposés aux agents infectieux sans que l'on sache s'ils sont infectés, ne permettaient pas de conclure si la mutation provoque une exagération de l'inflammation de l'hôte (et de ses conséquences), ou une sensibilité accrue à l'entrée et la multiplication de l'agent pathogène, ou les deux. Dans le [Chapitre III](#), nous avons aussi montré par des GWAS que cette mutation était associée à la quantité de Staphylocoques dans le lait. Ce résultat n'a pas été retrouvé dans l'analyse du challenge expérimental ([Chapitre IV](#)) où la variable de bactériologie n'était pas différente entre les C/C et les T/T. Ceci peut s'expliquer par les différentes situations au cours desquelles ont été mesurées ces variables, de la même manière que précédemment, c'est-à-dire qu'au cours de l'infection expérimentale, un mécanisme immunitaire se met en place de manière temporaire permettant l'élimination des bactéries tout

en gardant un phénomène inflammatoire présent chez les T/T. On peut faire l'hypothèse que la dégradation des tissus enflammés favorise de manière récurrente l'entrée de bactéries, d'où l'association sur la quantité de bactéries dans le lait chez des individus à l'état de « repos immunitaire » détectée par les GWAS. Le mode d'infection peut être également questionné, de même que la souche bactérienne choisie, qui ne reflètent peut-être pas tout à fait la situation naturelle.

Les infections expérimentales ([Chapitre IV.](#)) ont montré qu'une transcription de gènes en lien avec la défense immunitaire était activée à partir de 16 heures post-inoculation. Ces gènes activés permettraient la mise en place de mécanismes de prolifération, et/ou de différenciation, et/ou de migration des cellules immunitaires du sang vers le tissu mammaire infecté. Les différences obtenues entre les C/C et les T/T montrent que les populations de cellules qui semblent être les plus impactées par la mutation *SOCS2* sont les lymphocytes T, moins nombreux chez les mutants dans la mamelle infectée. Ceci suggère que la protéine SOCS-2 aurait un rôle dans la régulation de la réactivité des lymphocytes T au cours de la phase aiguë de l'infection. Cette différence aboutirait à un état d'inflammation maintenu chez les mutés après cette phase de recrutement, à T56 (fièvre et dégâts tissulaires), où on observe par ailleurs les gènes différenciellement exprimés. Les mécanismes sous-jacents mis en lumière dans ce travail de thèse sont les voies des interférons et de STAT-3. Cependant, des études complémentaires sont nécessaires afin de mieux caractériser les réactions en chaîne qui aboutissent à l'état clinique dégradé observé chez les mutants. Des études transcriptomiques unicellulaires pourraient être effectuées après challenge *in vitro* ou *in vivo* sur un modèle murin par exemple, en se focalisant sur les populations cellulaires d'intérêt (lymphocytes T CD4 et CD8).

Il est à noter que cette infection expérimentale pourrait être complétée avec des individus supplémentaires pour arriver au nombre minimal d'individus par groupe suggéré dans la saisine. Ce nombre était en effet de 11 brebis par groupe (afin d'observer un écart entre groupe significatif sur le SCS) et nous n'étions malheureusement qu'à sept dans nos analyses pour des raisons de disponibilité des brebis, et pour des raisons techniques. Ceci expliquerait le faible nombre de p-valeurs significatives observées pour ce caractère et justifie le seuil de 0,1 pour les p-valeurs retenues dans nos interprétations. Des phénotypes sont cependant plus impactés que le SCS puisque l'on observe des effets significatifs.

VI. Démonstration de la pléiotropie du gène *SOCS2* par les études génétiques

Dans le chapitre bibliographique ([Chapitre I.](#)), nous avons vu dans quelle mesure la protéine SOCS-2 avait un effet pléiotrope, c'est-à-dire jouant un rôle sur de nombreuses fonctions biologiques. Ceci est dû à sa nature de régulatrice de la grande voie de transmission de signal cellulaire qu'est la voie JAK/STAT. Nous avons ensuite affirmé l'association de la mutation R96C du gène *SOCS2* à différents caractères tels que la santé mammaire, la production laitière, la croissance, la reproduction, etc. ([Chapitre III.](#)), validée par ailleurs grâce à l'étude prospective sur la prise en compte de cette mutation dans les modèles d'évaluation génomique ([Chapitre II.](#)). Ces associations ont toujours été démontrées de manière indépendante les unes des autres dans des modèles uni-caractère.

Bien que non développée dans ce manuscrit, une approche multi-caractère a été appliquée au cours de cette thèse : la méthode CLIP test (David *et al.*, 2013), dont les résultats ont été présentés sous forme d'un poster associé à un papier court au congrès mondial de génétique animale appliquée aux animaux d'élevage (WCGALP) en février 2018 (Nouvelle-Zélande) (Annexe 22). Cette approche a confirmé que la région de *SOCS2* était bien pléiotrope et que les différents marqueurs de cette région n'étaient pas associés de manière différente aux différents caractères (effets corrélés). Les caractères testés étaient la production laitière, le SCS et la bactériologie par qPCR pour des brebis en première lactation, et le Poids Âge Type à 250 jours des mâles des stations de testage. De nombreuses autres approches sont développées aujourd'hui pour mettre en évidence la pléiotropie de la détermination génétique des caractères (Hackinger & Zeggini, 2017). Il serait intéressant de tester certaines de ces méthodes et notamment celles, appelées randomisation mendélienne, permettant de déterminer si la corrélation entre deux caractères est due à un lien causal (*i.e.* le caractère 1 est un facteur de risque causal pour le caractère 2) ou à des facteurs confusionnels comme l'exposition environnementale. En effet, à ce jour, concernant l'association de la mutation à la production laitière par exemple, deux hypothèses sont émises : lien direct *via* la régulation de la prolactine, ou effet indirect *via* la régulation de l'hormone de croissance qui joue sur le gabarit des brebis et ainsi sur leur capacité de production.

Références

- Abdoli, R., Zamani, P., Mirhoseini, S. Z., Hossein-Zadeh, N. G., Nadri, S.**, 2016. A review on prolificacy genes in sheep. *Reproduction in Domestic Animals*, 51, 5, 631-637. DOI: 10.1111/rda.12733.
- Adams, H. A., Sonstegard, T. S., VanRaden, P. M., Null, D. J., Van Tassell, C. P., Larkin, D. M., Lewin, H. A.**, 2016. Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science*, 99, 8, 6693-6701. DOI: 10.3168/jds.2015-10517.
- Alexander, W. S.**, 2002. Suppressors of cytokine signalling (SOCS) in the immune system. *Nature Reviews Immunology*, 2, 6, 410. DOI: 10.1038/nri818.
- Alexander, W. S., Starr, R., Fenner, J. E., Scott, C. L., Handman, E., Sprigg, N. S., Corbin, J. E., Cornish, A. L., Darwiche, R., Owczarek, C. M., Kay, T. W. H., Nicola, N. A., Hertzog, P. J., Metcalf, D., Hilton, D. J.**, 1999. SOCS1 Is a Critical Inhibitor of Interferon γ Signaling and Prevents the Potentially Fatal Neonatal Actions of this Cytokine. *Cell*, 98, 5, 597-608. DOI: 10.1016/S0092-8674(00)80047-1.
- Ali, A. K. A., Shook, G. E.**, 1980. An optimum transformation for somatic cell concentration in milk. *Journal of Dairy Science*, 63, 3, 487-490. DOI: 10.3168/jds.S0022-0302(80)82959-6.
- Aparecida Silva Barbosa, A., Freitas Versiani, A., Fonseca da Cunha Sousa, L., Silva de Miranda, A., Gasparini, M. R., Brant, F., Silva, D. G., Luisa Quintino-de-Carvalho, I., Marianetti Soriani, F., Guimarães da Fonseca, F., César Vasconcelos, A., da Silva Barcelos, L., Martins Teixeira, M., Lúcio Teixeira, A., Machado, F. S., Barbosa-Stancioli, E. F., Rachid, M. A.**, 2016. Role of the suppressor of cytokine signaling 2 (SOCS2) during meningoencephalitis caused by Bovine herpesvirus 5 (BoHV-5). *Comparative Immunology, Microbiology and Infectious Diseases*, 47, 26-31. DOI: 10.1016/j.cimid.2016.05.003.
- Argetsinger, L. S., Campbell, G. S., Yang, X., Witthuhn, B. A., Silvennoinen, O., Ihle, J. N., Carter-Su, C.**, 1993. Identification of JAK2 as a growth hormone receptor-associated tyrosine kinase. *Cell*, 74, 2, 237-244. DOI: 10.1016/0092-8674(93)90415-M.
- Astruc, J.-M., Baloche, G., Larroque, H., Beltran de Heredia, I., Labatut, J., Lagriffoul, G., Moreno, C., Robert-Granié, C., Boscher, M.-Y., Chantry-Darmon, C., Aguerre, X., Boulenc, P., Frégeat, G., Giral-Viala, B., Guibert, P., Panis, P., Soulas, C., Barillet, F., Legarra, A.**, 2012. La sélection génomique des ovins laitiers en France : stratégies, premiers résultats des évaluations génomiques et perspectives. *19èmes Rencontres autour des Recherches sur les Ruminants*. Paris (France), 81-84.
- Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.-M., Robert-Granié, C., Barillet, F.**, 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *Journal of Dairy Science*, 97, 2, 1107-1116. DOI: 10.3168/jds.2013-7135.
- Banks, A. S., Li, J., McKeag, L., Hribal, M. L., Kashiwada, M., Accili, D., Rothman, P. B.**, 2005. Deletion of SOCS7 leads to enhanced insulin action and enlarged islets of Langerhans. *The Journal of Clinical Investigation*, 115, 9, 2462-2471. DOI: 10.1172/JCI23853.

- Barillet, F., Lagriffoul, G., Marnet, P.-G., Larroque, H., Rupp, R., Portes, D., Bocquier, F., Astruc, J.-M.**, 2016. Objectifs de sélection et stratégie raisonnée de mise en oeuvre à l'échelle des populations de brebis laitières françaises. *INRA Productions Animales*, 29, 1, 19-40.
- Barwick, S. A., Henzell, A. L., Herd, R. M., Walmsley, B. J., Arthur, P. F.**, 2019. Methods and consequences of including reduction in greenhouse gas emission in beef cattle multiple-trait selection. *Genetics, selection, evolution: GSE*, 51, 1, 18. DOI: 10.1186/s12711-019-0459-5.
- Bates, D., Mächler, M., Bolker, B., Walker, S.**, 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1, 1-48. DOI: 10.18637/jss.v067.i01.
- Becker, D., Tetens, J., Brunner, A., Bürstel, D., Ganter, M., Kijas, J., Consortium, for the I. S. G., Drögemüller, C.**, 2010. Microphthalmia in Texel Sheep Is Associated with a Missense Mutation in the Paired-Like Homeodomain 3 (PITX3) Gene. *PLOS ONE*, 5, 1, e8689. DOI: 10.1371/journal.pone.0008689.
- Benavides, M. V., Sonstegard, T. S., Van Tassell, C.**, 2016. Genomic Regions Associated with Sheep Resistance to Gastrointestinal Nematodes. *Trends in Parasitology*, 32, 6, 470-480. DOI: 10.1016/j.pt.2016.03.007.
- Benjamini, Y., Hochberg, Y.**, 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 1, 289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.
- Bodin, L., Di Pasquale, E., Fabre, S., Bontoux, M., Monget, P., Persani, L., Mulsant, P.**, 2007. A novel mutation in the bone morphogenetic protein 15 gene causing defective protein secretion is associated with both increased ovulation rate and sterility in Lacaune sheep. *Endocrinology*, 148, 1, 393-400. DOI: 10.1210/en.2006-0764.
- Böhmer, F.-D., Friedrich, K.**, 2014. Protein tyrosine phosphatases as wardens of STAT signaling. *JAK-STAT*, 3, 1. DOI: 10.4161/jkst.28087.
- Boichard, D., Ducrocq, V., Croiseau, P., Fritz, S.**, 2016. Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies*, 339, 7, 274-277. DOI: 10.1016/j.crvi.2016.04.007.
- Boulanger, L., Pannetier, M., Gall, L., Allais-Bonnet, A., Elzaiat, M., Le Bourhis, D., Daniel, N., Richard, C., Cotinot, C., Ghyselinck, N. B., Pailhoux, E.**, 2014. FOXL2 Is a Female Sex-Determining Gene in the Goat. *Current Biology*, 24, 4, 404-408. DOI: 10.1016/j.cub.2013.12.039.
- Bourneuf, E., Otz, P., Pausch, H., Jagannathan, V., Michot, P., Grohs, C., Piton, G., Ammermüller, S., Deloche, M.-C., Fritz, S., Leclerc, H., Péchoux, C., Boukadiri, A., Hozé, C., Saintilan, R., Créchet, F., Mosca, M., Segelke, D., Guillaume, F., Bouet, S., Baur, A., Vasilescu, A., Genestout, L., Thomas, A., Allais-Bonnet, A., Rocha, D., Colle, M.-A., Klopp, C., Esquerré, D., Wurmser, C., Flisikowski, K., Schwarzenbacher, H., Burgstaller, J., Brüggmann, M., Dietschi, E., Rudolph, N., Freick, M., Barbey, S., Fayolle, G., Danchin-Burge, C., Schibler, L., Bed'Hom, B., Hayes, B. J., Daetwyler, H. D., Fries, R., Boichard, D., Pin, D., Drögemüller, C., Capitan, A.**, 2017. Rapid Discovery of De Novo Deleterious

Mutations in Cattle Enhances the Value of Livestock as Model Species. *Scientific Reports*, 7, 1, 11466. DOI: 10.1038/s41598-017-11523-3.

Boyle, K., Robb, L., 2008. The role of SOCS3 in modulating leukaemia inhibitory factor signalling during murine placental development. *Journal of Reproductive Immunology*, 77, 1, 1-6. DOI: 10.1016/j.jri.2007.02.003.

Bullock, A. N., Debreczeni, J. E., Edwards, A. M., Sundström, M., Knapp, S., 2006. Crystal structure of the SOCS2-elongin C-elongin B complex defines a prototypical SOCS box ubiquitin ligase. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 20, 7637-7642. DOI: 10.1073/pnas.0601638103.

Carillier-Jacquin, C., Larroque, H., Robert-Granié, C., 2016. Including α s1 casein gene information in genomic evaluations of French dairy goats. *Genetics Selection Evolution*, 48, 54. DOI: 10.1186/s12711-016-0233-x.

Caroli, A. M., Chessa, S., Erhardt, G. J., 2009. Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *Journal of Dairy Science*, 92, 11, 5335-5352. DOI: 10.3168/jds.2009-2461.

Carvalho, A. V., Reinaud, P., Forde, N., Healey, G. D., Eozenou, C., Giraud-Delville, C., Mansouri-Attia, N., Gall, L., Richard, C., Lonergan, P., Sheldon, I. M., Lea, R. G., Sandra, O., 2014. SOCS genes expression during physiological and perturbed implantation in bovine endometrium. *Reproduction (Cambridge, England)*, 148, 6, 545-557. DOI: 10.1530/REP-14-0214.

Charlier, C., Agerholm, J. S., Coppieters, W., Karlskov-Mortensen, P., Li, W., Jong, G. de, Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaart, E., Georges, M., Fredholm, M., 2012. A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. *PLOS ONE*, 7, 8, e43085. DOI: 10.1371/journal.pone.0043085.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B. R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R. J., Georges, M., 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics*, 40, 4, 449-454. DOI: 10.1038/ng.96.

Chen, Z., Laurence, A., Kanno, Y., Pacher-Zavisin, M., Zhu, B.-M., Tato, C., Yoshimura, A., Hennighausen, L., O'Shea, J. J., 2006. Selective regulatory function of Socs3 in the formation of IL-17-secreting T cells. *Proceedings of the National Academy of Sciences*, 103, 21, 8137-8142. DOI: 10.1073/pnas.0600666103.

Cheon, H., Yang, J., Stark, G. R., 2011. The Functions of Signal Transducers and Activators of Transcriptions 1 and 3 as Cytokine-Inducible Proteins. *Journal of Interferon & Cytokine Research*, 31, 1, 33-40. DOI: 10.1089/jir.2010.0100.

Chong, M. M. W., Cornish, A. L., Darwiche, R., Stanley, E. G., Purton, J. F., Godfrey, D. I., Hilton, D. J., Starr, R., Alexander, W. S., Kay, T. W. H., 2003. Suppressor of Cytokine

Signaling-1 Is a Critical Regulator of Interleukin-7-Dependent CD8⁺ T Cell Differentiation. *Immunity*, 18, 4, 475-487. DOI: 10.1016/S1074-7613(03)00078-5.

Chung, C. D., Liao, J., Liu, B., Rao, X., Jay, P., Berta, P., Shuai, K., 1997. Specific Inhibition of Stat3 Signal Transduction by PIAS3. *Science*, 278, 5344, 1803-1805. DOI: 10.1126/science.278.5344.1803.

Ciepluch, A., Rutkowska, K., Oprządek, J., Poławska, E., 2017. Genetic disorders in beef cattle: a review. *Genes & Genomics*, 39, 5, 461-471. DOI: 10.1007/s13258-017-0525-8.

Clark, E. L., Bush, S. J., McCulloch, M. E. B., Farquhar, I. L., Young, R., Lefevre, L., Pridans, C., Tsang, H. G., Wu, C., Afrasiabi, C., Watson, M., Whitelaw, C. B., Freeman, T. C., Summers, K. M., Archibald, A. L., Hume, D. A., 2017. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genetics*, 13, 9, e1006997. DOI: 10.1371/journal.pgen.1006997.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.-M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C., Georges, M., 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38, 7, 813. DOI: 10.1038/ng1810.

Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loor, J. J., Everts-van der Wind, A., Lee, J.-H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I., Ron, M., 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research*, 15, 7, 936-944. DOI: 10.1101/gr.3806705.

Collum, R. G., Brutsaert, S., Lee, G., Schindler, C., 2000. A Stat3-interacting protein (StIP1) regulates cytokine signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 18, 10120-10125.

Cornish, A. L., Davey, G. M., Metcalf, D., Purton, J. F., Corbin, J. E., Greenhalgh, C. J., Darwiche, R., Wu, L., Nicola, N. A., Godfrey, D. I., Heath, W. R., Hilton, D. J., Alexander, W. S., Starr, R., 2003. Suppressor of Cytokine Signaling-1 Has IFN- γ -Independent Actions in T Cell Homeostasis. *The Journal of Immunology*, 170, 2, 878-886. DOI: 10.4049/jimmunol.170.2.878.

Cosman, D., Lyman, S. D., Idzerda, R. L., Beckmann, M. P., Park, L. S., Goodwin, R. G., March, C. J., 1990. A new cytokine receptor superfamily. *Trends in Biochemical Sciences*, 15, 7, 265-270. DOI: 10.1016/0968-0004(90)90051-C.

Coux, O., Piechaczyk, M., 2000. Le système ubiquitine/protéasome : un ensemble (de) complexe(s) pour dégrader les protéines. *médecine/sciences*, 16, 5, 623. DOI: 10.4267/10608/1705.

Croker, B. A., Kiu, H., Nicholson, S. E., 2008. SOCS regulation of the JAK/STAT signalling pathway. *Seminars in Cell & Developmental Biology*, 19, 4, 414-422. DOI: 10.1016/j.semcdb.2008.07.010.

Crocker, B. A., Metcalf, D., Robb, L., Wei, W., Mifsud, S., DiRago, L., Cluse, L. A., Sutherland, K. D., Hartley, L., Williams, E., Zhang, J.-G., Hilton, D. J., Nicola, N. A., Alexander, W. S., Roberts, A. W., 2004. SOCS3 Is a Critical Physiological Negative Regulator of G-CSF Signaling and Emergency Granulopoiesis. *Immunity*, 20, 2, 153-165. DOI: 10.1016/S1074-7613(04)00022-6.

Davey, G. M., Starr, R., Cornish, A. L., Burghardt, J. T., Alexander, W. S., Carbone, F. R., Surh, C. D., Heath, W. R., 2005. SOCS-1 regulates IL-15–driven homeostatic proliferation of antigen-naïve CD8 T cells, limiting their autoimmune potential. *Journal of Experimental Medicine*, 202, 8, 1099-1108. DOI: 10.1084/jem.20050003.

David, I., Elsen, J.-M., Concordet, D., 2013. CLIP Test: a new fast, simple and powerful method to distinguish between linked or pleiotropic quantitative trait loci in linkage disequilibria analysis. *Heredity*, 110, 3, 232-238. DOI: 10.1038/hdy.2012.70.

Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., Martyniuk, E., Bodin, L., 2013. Genome-Wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperproliferacy Phenotype in Sheep. *PLOS Genetics*, 9, 4, e1003482. DOI: 10.1371/journal.pgen.1003482.

Deverrière, B., 2007. Reproduction expérimentale de mammites à *Staphylococcus aureus* chez la brebis : comparaison de lignées génétiques divergentes pour les comptages cellulaires. Université Paul Sabatier, Toulouse.

Dey, B. R., Spence, S. L., Nissley, P., Furlanetto, R. W., 1998. Interaction of Human Suppressor of Cytokine Signaling (SOCS)-2 with the Insulin-like Growth Factor-I Receptor. *Journal of Biological Chemistry*, 273, 37, 24095-24101. DOI: 10.1074/jbc.273.37.24095.

Dobie, R., Ahmed, S. F., Staines, K. A., Pass, C., Jasim, S., MacRae, V. E., Farquharson, C., 2015. Increased linear bone growth by GH in the absence of SOCS2 is independent of IGF-1. *Journal of Cellular Physiology*, 230, 11, 2796-2806. DOI: 10.1002/jcp.25006.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T. R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 1, 15-21. DOI: 10.1093/bioinformatics/bts635.

Drouilhet, L., Mansanet, C., Sarry, J., Tabet, K., Bardou, P., Woloszyn, F., Lluch, J., Harichaux, G., Viguié, C., Monniaux, D., Bodin, L., Mulsant, P., Fabre, S., 2013. The highly prolific phenotype of Lacaune sheep is associated with an ectopic expression of the B4GALNT2 gene within the ovary. *PLoS genetics*, 9, 9, e1003809. DOI: 10.1371/journal.pgen.1003809.

Duchemin, S. I., Colombani, C., Legarra, A., Baloché, G., Larroque, H., Astruc, J.-M., Barillet, F., Robert-Granié, C., Manfredi, E., 2012. Genomic selection in the French Lacaune dairy sheep breed. *Journal of Dairy Science*, 95, 5, 2723-2733. DOI: 10.3168/jds.2011-4980.

Eckersall, P. D., Bell, R., 2010. Acute phase proteins: Biomarkers of infection and inflammation in veterinary medicine. *The Veterinary Journal*, 185, 1, 23-27. DOI: 10.1016/j.tvjl.2010.04.009.

- Emilsson, V., Arch, J. R., de Groot, R. P., Lister, C. A., Cawthorne, M. A.,** 1999. Leptin treatment increases suppressors of cytokine signaling in central and peripheral tissues. *FEBS letters*, 455, 1-2, 170-174.
- Endo, T. A., Masuhara, M., Yokouchi, M., Suzuki, R., Sakamoto, H., Mitsui, K., Matsumoto, A., Tanimura, S., Ohtsubo, M., Misawa, H., Miyazaki, T., Leonor, N., Taniguchi, T., Fujita, T., Kanakura, Y., Komiya, S., Yoshimura, A.,** 1997. A new protein containing an SH2 domain that inhibits JAK kinases. *Nature*, 387, 6636, 921-924. DOI: 10.1038/43213.
- Endo, T., Sasaki, A., Minoguchi, M., Joo, A., Yoshimura, A.,** 2003. CIS1 Interacts with the Y532 of the Prolactin Receptor and Suppresses Prolactin-Dependent STAT5 Activation. *The Journal of Biochemistry*, 133, 1, 109-113. DOI: 10.1093/jb/mvg004.
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A., Goddard, M. E.,** 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95, 7, 4114-4129. DOI: 10.3168/jds.2011-5019.
- Eyckerman, S., Verhee, A., Heyden, J. V. der, Lemmens, I., Ostade, X. V., Vandekerckhove, J., Tavernier, J.,** 2001. Design and application of a cytokine-receptor-based interaction trap. *Nature Cell Biology*, 3, 12, 1114. DOI: 10.1038/ncb1201-1114.
- Farabegoli, F., Ceccarelli, C., Santini, D., Taffurelli, M.,** 2005. Suppressor of cytokine signalling 2 (SOCS-2) expression in breast carcinoma. *Journal of Clinical Pathology*, 58, 10, 1046-1050. DOI: 10.1136/jcp.2004.024919.
- Fasquelle, C., Sartelet, A., Li, W., Dive, M., Tamma, N., Michaux, C., Druet, T., Huijbers, I. J., Isacke, C. M., Coppieters, W., Georges, M., Charlier, C.,** 2009. Balancing Selection of a Frame-Shift Mutation in the MRC2 Gene Accounts for the Outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLOS Genetics*, 5, 9, e1000666. DOI: 10.1371/journal.pgen.1000666.
- Favre, H., Benhamou, A., Finidori, J., Kelly, P. A., Edery, M.,** 1999. Dual effects of suppressor of cytokine signaling (SOCS-2) on growth hormone signal transduction. *FEBS Letters*, 453, 1, 63-66. DOI: 10.1016/S0014-5793(99)00681-X.
- Fernando, R. L., Dekkers, J. C., Garrick, D. J.,** 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution*, 46, 1, 50. DOI: 10.1186/1297-9686-46-50.
- Ferrao, R., Lupardus, P. J.,** 2017. The Janus Kinase (JAK) FERM and SH2 Domains: Bringing Specificity to JAK-Receptor Interactions. *Frontiers in Endocrinology*, 8. DOI: 10.3389/fendo.2017.00071.
- Firmbach-Kraft, I., Byers, M., Shows, T., Dalla-Favera, R., Krolewski, J. J.,** 1990. tyk2, prototype of a novel class of non-receptor tyrosine kinase genes. *Oncogene*, 5, 9, 1329-1336.
- Fisher, R. A.,** 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399-433.

Fitzgerald, J. S., Toth, B., Jeschke, U., Schleussner, E., Markert, U. R., 2009. Knocking off the suppressors of cytokine signaling (SOCS): their roles in mammalian pregnancy. *Journal of Reproductive Immunology*, 83, 1, 117-123. DOI: 10.1016/j.jri.2009.07.010.

Fritz, S., Capitan, A., Djari, A., Rodriguez, S. C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D., Boichard, D., 2013. Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PLOS ONE*, 8, 6, e65550. DOI: 10.1371/journal.pone.0065550.

Fritz, S., Hoze, C., Rebours, E., Barbat, A., Bizard, M., Chamberlain, A., Escouflaire, C., Vander Jagt, C., Boussaha, M., Grohs, C., Allais-Bonnet, A., Philippe, M., Vallée, A., Amigues, Y., Hayes, B. J., Boichard, D., Capitan, A., 2018. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science*, 101, 7, 6220-6231. DOI: 10.3168/jds.2017-14119.

Fujimoto, M., Naka, T., 2003. Regulation of cytokine signaling by SOCS family molecules. *Trends in Immunology*, 24, 12, 659-666. DOI: 10.1016/j.it.2003.10.008.

Fujimoto, M., Tsutsui, H., Yumikura-Futatsugi, S., Ueda, H., Xingshou, O., Abe, T., Kawase, I., Nakanishi, K., Kishimoto, T., Naka, T., 2002. A regulatory role for suppressor of cytokine signaling-1 in Th polarization in vivo. *International Immunology*, 14, 11, 1343-1350. DOI: 10.1093/intimm/dxf094.

Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P. E., Juengel, J. L., Jokiranta, T. S., McLaren, R. J., Luiro, K., Dodds, K. G., Montgomery, G. W., Beattie, A. E., Davis, G. H., Ritvos, O., 2000. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics*, 25, 3, 279. DOI: 10.1038/77033.

Gao, H., Koivula, M., Jensen, J., Strandén, I., Madsen, P., Pitkänen, T., Aamand, G. P., Mäntysaari, E. A., 2018. Short communication: Genomic prediction using different single-step methods in the Finnish red dairy cattle population. *Journal of Dairy Science*, 101, 11, 10082-10088. DOI: 10.3168/jds.2018-14913.

Ghoreschi, K., Laurence, A., O’Shea, J. J., 2009. Janus kinases in immune cell signaling. *Immunological reviews*, 228, 1, 273-287. DOI: 10.1111/j.1600-065X.2008.00754.x.

Goddard, M. E., Kemper, K. E., MacLeod, M. I., Chamberlain, A. J., Hayes, B. J., 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proceedings of the Royal Society B: Biological Sciences*, 283, 1835, 20160569. DOI: 10.1098/rspb.2016.0569.

Goldmann, W., Hunter, N., Foster, J. D., Salbaum, J. M., Beyreuther, K., Hope, J., 1990. Two alleles of a neural protein gene linked to scrapie in sheep. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 7, 2476-2480. DOI: 10.1073/pnas.87.7.2476.

Goldshmit, Y., Walters, C. E., Scott, H. J., Greenhalgh, C. J., Turnley, A. M., 2004. SOCS2 Induces Neurite Outgrowth by Regulation of Epidermal Growth Factor Receptor Activation. *Journal of Biological Chemistry*, 279, 16, 16349-16355. DOI: 10.1074/jbc.M312873200.

- Gossner, A., Wilkie, H., Joshi, A., Hopkins, J.,** 2013. Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. *Veterinary Research*, 44, 1, 68. DOI: 10.1186/1297-9716-44-68.
- Greenhalgh, C. J., Bertolino, P., Asa, S. L., Metcalf, D., Corbin, J. E., Adams, T. E., Davey, H. W., Nicola, N. A., Hilton, D. J., Alexander, W. S.,** 2002a. Growth Enhancement in Suppressor of Cytokine Signaling 2 (SOCS-2)-Deficient Mice Is Dependent on Signal Transducer and Activator of Transcription 5b (STAT5b). *Molecular Endocrinology*, 16, 6, 1394-1406. DOI: 10.1210/mend.16.6.0845.
- Greenhalgh, C. J., Hilton, D. J.,** 2001. Negative regulation of cytokine signaling. *Journal of Leukocyte Biology*, 70, 3, 348-356. DOI: 10.1189/jlb.70.3.348.
- Greenhalgh, C. J., Metcalf, D., Thaus, A. L., Corbin, J. E., Uren, R., Morgan, P. O., Fabri, L. J., Zhang, J.-G., Martin, H. M., Willson, T. A., Billestrup, N., Nicola, N. A., Baca, M., Alexander, W. S., Hilton, D. J.,** 2002b. Biological Evidence That SOCS-2 Can Act Either as an Enhancer or Suppressor of Growth Hormone Signaling. *Journal of Biological Chemistry*, 277, 43, 40181-40184. DOI: 10.1074/jbc.C200450200.
- Greenhalgh, C. J., Rico-Bautista, E., Lorentzon, M., Thaus, A. L., Morgan, P. O., Willson, T. A., Zervoudakis, P., Metcalf, D., Street, I., Nicola, N. A., Nash, A. D., Fabri, L. J., Norstedt, G., Ohlsson, C., Flores-Morales, A., Alexander, W. S., Hilton, D. J.,** 2005. SOCS2 negatively regulates growth hormone action in vitro and in vivo. *Journal of Clinical Investigation*, 115, 2, 397-406. DOI: 10.1172/JCI200522710.
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., Snell, R.,** 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. *Genome Research*, 12, 2, 222-231. DOI: 10.1101/gr.224202.
- Grobet, L., Martin, L. J. R., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Ménessier, F., Massabanda, J., Fries, R., Hanset, R., Georges, M.,** 1997. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature Genetics*, 17, 1, 71. DOI: 10.1038/ng0997-71.
- Hackinger, S., Zeggini, E.,** 2017. Statistical methods to detect pleiotropy in human complex traits. *Open Biology*, 7, 11. DOI: 10.1098/rsob.170125.
- Haffner, M. C., Petridou, B., Peyrat, J. P., Révillion, F., Müller-Holzner, E., Daxenbichler, G., Marth, C., Doppler, W.,** 2007. Favorable prognostic value of SOCS2 and IGF-I in breast cancer. *BMC Cancer*, 7, 1, 136. DOI: 10.1186/1471-2407-7-136.
- Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R., Galloway, S. M.,** 2004. Mutations in the Genes for Oocyte-Derived Growth Factors GDF9 and BMP15 Are Associated with Both Increased Ovulation Rate and Sterility in Cambridge and Belclare Sheep (*Ovis aries*). *Biology of Reproduction*, 70, 4, 900-909. DOI: 10.1095/biolreprod.103.023093.

- Hansen, J. A., Lindberg, K., Hilton, D. J., Nielsen, J. H., Billestrup, N.**, 1999. Mechanism of Inhibition of Growth Hormone Receptor Signaling by Suppressor of Cytokine Signaling Proteins. *Molecular Endocrinology*, 13, 11, 1832-1843. DOI: 10.1210/mend.13.11.0368.
- Harris, J., Stanford, P. M., Sutherland, K., Oakes, S. R., Naylor, M. J., Robertson, F. G., Blazek, K. D., Kazlauskas, M., Hilton, H. N., Wittlin, S., Alexander, W. S., Lindeman, G. J., Visvader, J. E., Ormandy, C. J.**, 2006. Socs2 and elf5 mediate prolactin-induced mammary gland development. *Molecular Endocrinology (Baltimore, Md.)*, 20, 5, 1177-1187. DOI: 10.1210/me.2005-0473.
- Hayes, B. J., Donoghue, K. A., Reich, C. M., Mason, B. A., Bird-Gardiner, T., Herd, R. M., Arthur, P. F.**, 2016. Genomic heritabilities and genomic estimated breeding values for methane traits in Angus cattle. *Journal of Animal Science*, 94, 3, 902-908. DOI: 10.2527/jas.2015-0078.
- Hazel, L. N.**, 1943. The Genetic Basis for Constructing Selection Indexes. *Genetics*, 28, 6, 476-490.
- Heaton, M. P., Clawson, M. L., Chitko-Mckown, C. G., Leymaster, K. A., Smith, T. P. L., Harhay, G. P., White, S. N., Herrmann-Hoesing, L. M., Mousel, M. R., Lewis, G. S., Kalbfleisch, T. S., Keen, J. E., Laegreid, W. W.**, 2012. Reduced Lentivirus Susceptibility in Sheep with TMEM154 Mutations. *PLOS Genetics*, 8, 1. DOI: 10.1371/journal.pgen.1002467.
- Henderson, C. R.**, 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, 31, 2, 423-447. DOI: 10.2307/2529430.
- Hilton, D. J., Richardson, R. T., Alexander, W. S., Viney, E. M., Willson, T. A., Sprigg, N. S., Starr, R., Nicholson, S. E., Metcalf, D., Nicola, N. A.**, 1998. Twenty proteins containing a C-terminal SOCS box form five structural classes. *Proceedings of the National Academy of Sciences*, 95, 1, 114-119. DOI: 10.1073/pnas.95.1.114.
- Horvat, S., Medrano, J. F.**, 2001. Lack of Socs2 Expression Causes the High-Growth Phenotype in Mice. *Genomics*, 72, 2, 209-212. DOI: 10.1006/geno.2000.6441.
- Hoze, C., Fritz, S., Baur, A., Grohs, C., Danchin-Burge, C., Boichard, D.**, 2018. Prise en compte des gènes d'intérêt dans les objectifs de sélection en bovins (...) - 3R - Rencontres autour des Recherches sur les Ruminants. Paris (France).
- Hu, J., Winqvist, O., Flores-Morales, A., Wikström, A.-C., Norstedt, G.**, 2009. SOCS2 influences LPS induced human monocyte-derived dendritic cell maturation. *PloS One*, 4, 9, e7178. DOI: 10.1371/journal.pone.0007178.
- Igaz, P., Tóth, S., Falus, A.**, 2001. Biological and clinical significance of the JAK-STAT pathway; lessons from knockout mice. *Inflammation Research*, 50, 9, 435-441. DOI: 10.1007/PL00000267.
- Ivashkiv, L. B., Hu, X.**, 2004. Signaling by STATs. *Arthritis Research & Therapy*, 6, 4, 159-168. DOI: 10.1186/ar1197.
- Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Nielsen, U. S., Panitz, F., Aamand, G. P., Schulman, N., Georges, M., Vilkki, J., Lund, M.**

- S., Druet, T.**, 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLOS Genetics*, 10, 1, e1004049. DOI: 10.1371/journal.pgen.1004049.
- Kamalludin, M. H., Garcia-Guerra, A., Wiltbank, M. C., Kirkpatrick, B. W.**, 2018. Trio, a novel high fecundity allele: I. Transcriptome analysis of granulosa cells from carriers and noncarriers of a major gene for bovine ovulation rate. *Biology of Reproduction*, 98, 3, 323-334. DOI: 10.1093/biolre/iox133.
- Kambadur, R., Sharma, M., Smith, T. P. L., Bass, J. J.**, 1997. Mutations in myostatin (GDF8) in Double-Muscled Belgian Blue and Piedmontese Cattle. *Genome Research*, 7, 9, 910-915. DOI: 10.1101/gr.7.9.910.
- Kennedy, B. W., Quinton, M., van Arendonk, J. a. M.**, 1992. Estimation of effects of single genes on quantitative traits. *Journal of Animal Science*, 70, 7, 2000-2012. DOI: 10.2527/1992.7072000x.
- Kessler, D. S., Levy, D. E., Darnell, J. E.**, 1988a. Two interferon-induced nuclear factors bind a single promoter element in interferon-stimulated genes. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 22, 8521-8525. DOI: 10.1073/pnas.85.22.8521.
- Kessler, D. S., Pine, R., Pfeffer, L. M., Levy, D. E., Darnell, J. E.**, 1988b. Cells resistant to interferon are defective in activation of a promoter-binding factor. *The EMBO Journal*, 7, 12, 3779-3783.
- Kim, T. K., Maniatis, T.**, 1996. Regulation of Interferon- γ -Activated STAT1 by the Ubiquitin-Proteasome Pathway. *Science*, 273, 5282, 1717-1719. DOI: 10.1126/science.273.5282.1717.
- Kim, W. S., Kim, M. J., Kim, D. O., Byun, J.-E., Huy, H., Song, H. Y., Park, Y.-J., Kim, T.-D., Yoon, S. R., Choi, E.-J., Jung, H., Choi, I.**, 2017. Suppressor of Cytokine Signaling 2 Negatively Regulates NK Cell Differentiation by Inhibiting JAK2 Activity. *Scientific Reports*, 7, 46153. DOI: 10.1038/srep46153.
- Kipp, S., Segelke, D., Schierenbeck, S., Reinhardt, F., Reents, R., Wurmser, C., Pausch, H., Fries, R., Thaller, G., Tetens, J., Pott, J., Haas, D., Raddatz, B. B., Hewicker-Trautwein, M., Proios, I., Schmicke, M., Grünberg, W.**, 2016. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. *Journal of Dairy Science*, 99, 11, 8915-8931. DOI: 10.3168/jds.2016-11118.
- Kishimoto, T., Taga, T., Akira, S.**, 1994. Cytokine signal transduction. *Cell*, 76, 2, 253-262. DOI: 10.1016/0092-8674(94)90333-6.
- Knosp, C. A., Carroll, H. P., Elliott, J., Saunders, S. P., Nel, H. J., Amu, S., Pratt, J. C., Spence, S., Doran, E., Cooke, N., Jackson, R., Swift, J., Fitzgerald, D. C., Heaney, L. G., Fallon, P. G., Kissenpfennig, A., Johnston, J. A.**, 2011. SOCS2 regulates T helper type 2 differentiation and the generation of type 2 allergic responses. *The Journal of Experimental Medicine*, 208, 7, 1523-1531. DOI: 10.1084/jem.20101167.

- Kohanbash, G., Okada, H.**, 2012. MicroRNAs and STAT interplay. *Seminars in Cancer Biology*, 22, 1, 70-75. DOI: 10.1016/j.semcancer.2011.12.010.
- Krämer, A., Green, J., Pollard, J., Tugendreich, S.**, 2014. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics (Oxford, England)*, 30, 4, 523-530. DOI: 10.1093/bioinformatics/btt703.
- Krebs, D. L., Uren, R. T., Metcalf, D., Rakar, S., Zhang, J.-G., Starr, R., Souza, D. P. D., Hanzinikolas, K., Eyles, J., Connolly, L. M., Simpson, R. J., Nicola, N. A., Nicholson, S. E., Baca, M., Hilton, D. J., Alexander, W. S.**, 2002. SOCS-6 Binds to Insulin Receptor Substrate 4, and Mice Lacking the SOCS-6 Gene Exhibit Mild Growth Retardation. *Molecular and Cellular Biology*, 22, 13, 4567-4578. DOI: 10.1128/MCB.22.13.4567-4578.2002.
- Kung, W.-W., Ramachandran, S., Makukhin, N., Bruno, E., Ciulli, A.**, 2019. Structural insights into substrate recognition by the SOCS2 E3 ubiquitin ligase. *Nature Communications*, 10, 1, 2534. DOI: 10.1038/s41467-019-10190-4.
- Lande, R., Thompson, R.**, 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124, 3, 743-756.
- Lang, R., Pauleau, A.-L., Parganas, E., Takahashi, Y., Mages, J., Ihle, J. N., Rutschman, R., Murray, P. J.**, 2003. SOCS3 regulates the plasticity of gp130 signaling. *Nature Immunology*, 4, 6, 546. DOI: 10.1038/ni932.
- Laplanche, J. L., Chatelain, J., Westaway, D., Thomas, S., Dussaucy, M., Brugere-Picoux, J., Launay, J. M.**, 1993. PrP Polymorphisms Associated with Natural Scrapie Discovered by Denaturing Gradient Gel Electrophoresis. *Genomics*, 15, 1, 30-37. DOI: 10.1006/geno.1993.1006.
- Lassoued, N., Benkhilil, Z., Woloszyn, F., Rejeb, A., Aouina, M., Rekik, M., Fabre, S., Bedhiaf-Romdhani, S.**, 2017. FecX Bar a Novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. *BMC genetics*, 18, 1, 43. DOI: 10.1186/s12863-017-0510-x.
- Lavens, D., Montoye, T., Piessevaux, J., Zabeau, L., Vandekerckhove, J., Gevaert, K., Becker, W., Eyckerman, S., Tavernier, J.**, 2006. A complex interaction pattern of CIS and SOCS2 with the leptin receptor. *Journal of Cell Science*, 119, 11, 2214-2224. DOI: 10.1242/jcs.02947.
- Lê Cao, K.-A., González, I., Déjean, S.**, 2009. integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics (Oxford, England)*, 25, 21, 2855-2856. DOI: 10.1093/bioinformatics/btp515.
- Lebrun, P., Cognard, E., Gontard, P., Bellon-Paul, R., Filloux, C., Berthault, M. F., Magnan, C., Ruberte, J., Luppò, M., Pujol, A., Pachera, N., Herchuelz, A., Bosch, F., Van Obberghen, E.**, 2010. The suppressor of cytokine signalling 2 (SOCS2) is a key repressor of insulin secretion. *Diabetologia*, 53, 9, 1935-1946. DOI: 10.1007/s00125-010-1786-9.
- Legarra, A., Vitezica, Z. G.**, 2015. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. *Genetics Selection Evolution*, 47, 89. DOI: 10.1186/s12711-015-0165-x.

- Lenth, R. V.**, 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69, 1, 1-33. DOI: 10.18637/jss.v069.i01.
- Letellier, E., Haan, S.**, 2016. SOCS2: physiological and pathological functions. *Frontiers in Bioscience (Elite Edition)*, 8, 189-204.
- Levy, D. E., Darnell, J. E.**, 2002. STATs: transcriptional control and biological impact. *Nature Reviews Molecular Cell Biology*, 3, 9, 651. DOI: 10.1038/nrm909.
- Li, B., Dewey, C. N.**, 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 1, 323. DOI: 10.1186/1471-2105-12-323.
- Li, S., Chen, S., Xu, X., Sundstedt, A., Paulsson, K. M., Anderson, P., Karlsson, S., Sjögren, H.-O., Wang, P.**, 2000. Cytokine-Induced Src Homology 2 Protein (Cis) Promotes T Cell Receptor-Mediated Proliferation and Prolongs Survival of Activated T Cells. *Journal of Experimental Medicine*, 191, 6, 985-994. DOI: 10.1084/jem.191.6.985.
- Lindhé, B., Philipsson, J.**, 1998. Conventional breeding programmes and genetic resistance to animal diseases. *Revue Scientifique Et Technique (International Office of Epizootics)*, 17, 1, 291-301.
- Lohi, O., Lehto, V.-P.**, 2001. STAM/EAST/Hbp adapter proteins – integrators of signalling pathways. *FEBS Letters*, 508, 3, 287-290. DOI: 10.1016/S0014-5793(01)03079-4.
- Ma, L., Cole, J. B., Da, Y., VanRaden, P. M.**, 2019. Symposium review: Genetics, genome-wide association study, and genetic improvement of dairy fertility traits*. *Journal of Dairy Science*, 102, 4, 3735-3743. DOI: 10.3168/jds.2018-15269.
- Machado, F. S., Johndrow, J. E., Esper, L., Dias, A., Bafica, A., Serhan, C. N., Aliberti, J.**, 2006. Anti-inflammatory actions of lipoxin A4 and aspirin-triggered lipoxin are SOCS-2 dependent. *Nature Medicine*, 12, 3, 330-334. DOI: 10.1038/nm1355.
- Mäntysaari, E., Van Vleck, L. D.**, 1989. Estimation of genetic parameters for production and reproduction in Finnish Ayrshire cattle. *Journal of Dairy Science*, 72, 9, 2375-2386. DOI: 10.3168/jds.S0022-0302(89)79370-X.
- Marletta, D., Criscione, A., Bordonaro, S., Guastella, A. M., D’Urso, G.**, 2007. Casein polymorphism in goat’s milk. *Le Lait*, 87, 6, 491-504. DOI: 10.1051/lait:2007034.
- Martal, J. L., Chêne, N. M., Huynh, L. P., L’Haridon, R. M., Reinaud, P. B., Guillomot, M. W., Charlier, M. A., Charpigny, S. Y.**, 1998. IFN-tau: a novel subtype I IFN1. Structural characteristics, non-ubiquitous expression, structure-function relationships, a pregnancy hormonal embryonic signal and cross-species therapeutic potentialities. *Biochimie*, 80, 8-9, 755-777. DOI: 10.1016/S0300-9084(99)80029-7.
- Martin, M.**, 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, 1, 10-12. DOI: 10.14806/ej.17.1.200.

- Martin, P., Palhière, I., Maroteau, C., Bardou, P., Canale-Tabet, K., Sarry, J., Woloszyn, F., Bertrand-Michel, J., Racke, I., Besir, H., Rupp, R., Tosser-Klopp, G., 2017.** A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Scientific Reports*, 7, 1, 1872. DOI: 10.1038/s41598-017-02052-0.
- Martinez-Royo, A., Jurado, J. J., Smulders, J. P., Martí, J. I., Alabart, J. L., Roche, A., Fantova, E., Bodin, L., Mulsant, P., Serrano, M., Folch, J., Calvo, J. H., 2008.** A deletion in the bone morphogenetic protein 15 gene causes sterility and increased prolificacy in Rasa Aragonesa sheep. *Animal Genetics*, 39, 3, 294-297. DOI: 10.1111/j.1365-2052.2008.01707.x.
- Matsumoto, A., Masuhara, M., Mitsui, K., Yokouchi, M., Ohtsubo, M., Misawa, H., Miyajima, A., Yoshimura, A., 1997.** CIS, a Cytokine Inducible SH2 Protein, Is a Target of the JAK-STAT5 Pathway and Modulates STAT5 Activation. *Blood*, 89, 9, 3148-3154.
- Matsumoto, A., Seki, Y., Kubo, M., Ohtsuka, S., Suzuki, A., Hayashi, I., Tsuji, K., Nakahata, T., Okabe, M., Yamada, S., Yoshimura, A., 1999.** Suppression of STAT5 Functions in Liver, Mammary Glands, and T Cells in Cytokine-Inducible SH2-Containing Protein 1 Transgenic Mice. *Molecular and Cellular Biology*, 19, 9, 6396-6407. DOI: 10.1128/MCB.19.9.6396.
- Maures, T. J., Kurzer, J. H., Carter-Su, C., 2007.** SH2B1 (SH2-B) and JAK2: a multifunctional adaptor protein and kinase made for each other. *Trends in Endocrinology & Metabolism*, 18, 1, 38-45. DOI: 10.1016/j.tem.2006.11.007.
- McLaren, A., Kaseja, K., Yates, J., Mucha, S., Lambe, N. R., Conington, J., 2018.** New mastitis phenotypes suitable for genomic selection in meat sheep and their genetic relationships with udder conformation and lamb live weights. *Animal*, 1-10. DOI: 10.1017/S1751731118000393.
- Mense, K., Meyerholz, M., Gil Araujo, M., Lietzau, M., Knaack, H., Wrenzycki, C., Hoedemaker, M., Piechotta, M., 2015.** The somatotropic axis during the physiological estrus cycle in dairy heifers--Effect on hepatic expression of GHR and SOCS2. *Journal of Dairy Science*, 98, 4, 2409-2418. DOI: 10.3168/jds.2014-8734.
- Mesbah-Uddin, M., Hoze, C., Michot, P., Barbat, A., Lefebvre, R., Boussaha, M., Sahana, G., Fritz, S., Boichard, D., Capitan, A., 2019.** A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. *Journal of Dairy Science*, 102, 7, 6340-6356. DOI: 10.3168/jds.2018-16100.
- Metcalf, D., Greenhalgh, C. J., Viney, E., Willson, T. A., Starr, R., Nicola, N. A., Hilton, D. J., Alexander, W. S., 2000.** Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature*, 405, 6790, 1069-1073. DOI: 10.1038/35016611.
- Meuwissen, T. H., Hayes, B. J., Goddard, M. E., 2001.** Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 4, 1819-1829.
- Mohr, A., Chatain, N., Domoszlai, T., Rinis, N., Sommerauer, M., Vogt, M., Müller-Newen, G., 2012.** Dynamics and non-canonical aspects of JAK/STAT signalling. *European Journal of Cell Biology*, 91, 6, 524-532. DOI: 10.1016/j.ejcb.2011.09.005.

Mori, H., Hanada, R., Hanada, T., Aki, D., Mashima, R., Nishinakamura, H., Torisu, T., Chien, K. R., Yasukawa, H., Yoshimura, A., 2004. Socs3 deficiency in the brain elevates leptin sensitivity and confers resistance to diet-induced obesity. *Nature Medicine*, 10, 7, 739. DOI: 10.1038/nm1071.

Mullen, M. P., Hanrahan, J. P., 2014. Direct evidence on the contribution of a missense mutation in GDF9 to variation in ovulation rate of Finnsheep. *PloS One*, 9, 4, e95251. DOI: 10.1371/journal.pone.0095251.

Mulsant, P., Lecerf, F., Fabre, S., Schibler, L., Monget, P., Lanneluc, I., Pisselet, C., Riquet, J., Monniaux, D., Callebaut, I., Cribiu, E., Thimonier, J., Teyssier, J., Bodin, L., Cognié, Y., Chitour, N., Elsen, J.-M., 2001. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. *Proceedings of the National Academy of Sciences*, 98, 9, 5104-5109. DOI: 10.1073/pnas.091577598.

Murray, P. J., 2007. The JAK-STAT Signaling Pathway: Input and Output Integration. *The Journal of Immunology*, 178, 5, 2623-2629. DOI: 10.4049/jimmunol.178.5.2623.

Naka, T., Narazaki, M., Hirata, M., Matsumoto, T., Minamoto, S., Aono, A., Nishimoto, N., Kajita, T., Taga, T., Yoshizaki, K., Akira, S., Kishimoto, T., 1997. Structure and function of a new STAT-induced STAT inhibitor. *Nature*, 387, 6636, 924-929. DOI: 10.1038/43219.

Nakagawa, R., Naka, T., Tsutsui, H., Fujimoto, M., Kimura, A., Abe, T., Seki, E., Sato, S., Takeuchi, O., Takeda, K., Akira, S., Yamanishi, K., Kawase, I., Nakanishi, K., Kishimoto, T., 2002. SOCS-1 Participates in Negative Regulation of LPS Responses. *Immunity*, 17, 5, 677-687. DOI: 10.1016/S1074-7613(02)00449-1.

Ndiaye, K., Castonguay, A., Benoit, G., Silversides, D. W., Lussier, J. G., 2016. Differential regulation of Janus kinase 3 (JAK3) in bovine preovulatory follicles and identification of JAK3 interacting proteins in granulosa cells. *Journal of Ovarian Research*, 9, 1, 71. DOI: 10.1186/s13048-016-0280-5.

Nicol, L., Bishop, S. C., Pong-Wong, R., Bendixen, C., Holm, L.-E., Rhind, S. M., McNeilly, A. S., 2009. Homozygosity for a single base-pair mutation in the oocyte-specific GDF9 gene results in sterility in Thoka sheep. *Reproduction*, 138, 6, 921-933. DOI: 10.1530/REP-09-0193.

Nielsen, S. S., Denwood, M. J., Forkman, B., Houe, H., 2017. Selection of Meat Inspection Data for an Animal Welfare Index in Cattle and Pigs in Denmark. *Animals: an open access journal from MDPI*, 7, 12. DOI: 10.3390/ani7120094.

O'Brien, A. C., McHugh, N., Wall, E., Pabiou, T., McDermott, K., Randles, S., Fair, S., Berry, D. P., 2017. Genetic parameters for lameness, mastitis and dagginess in a multi-breed sheep population. *Animal*, 11, 6, 911-919. DOI: 10.1017/S1751731116002445.

O'Brien, K. B., O'Shea, J. J., Carter-Su, C., 2002. SH2-B Family Members Differentially Regulate JAK Family Tyrosine Kinases. *Journal of Biological Chemistry*, 277, 10, 8673-8681. DOI: 10.1074/jbc.M109165200.

Oget, C., Allain, C., Portes, D., Foucras, G., Stella, A., Astruc, J.-M., Sarry, J., Tossier-Klopp, G., Rupp, R., 2019a. A validation study of loci associated with mastitis resistance in

two French dairy sheep breeds. *Genetics Selection Evolution*, 51, 1, 5. DOI: 10.1186/s12711-019-0448-8.

Oget, C., Teissier, M., Astruc, J.-M., Tosser-Klopp, G., Rupp, R., 2019b. Alternative methods improve the accuracy of genomic prediction using information from a causal point mutation in a dairy sheep model. *BMC Genomics*, 20, 1, 719. DOI: 10.1186/s12864-019-6068-4.

Oget, C., Tosser-Klopp, G., Rupp, R., 2019c. Genetic and genomic studies in ovine mastitis. *Small Ruminant Research*, 176, 55-64. DOI: 10.1016/j.smallrumres.2019.05.011.

Pailhoux, E., Vigier, B., Chaffaux, S., Servel, N., Taourit, S., Furet, J. P., Fellous, M., Grosclaude, F., Cribiu, E. P., Cotinot, C., Vaiman, D., 2001. A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature Genetics*, 29, 4, 453-458. DOI: 10.1038/ng769.

Pailhoux, E., Vigier, B., Vaiman, D., Servel, N., Chaffaux, S., Cribiu, E. P., Cotinot, C., 2002. Ontogenesis of female-to-male sex-reversal in XX polled goats. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 224, 1, 39-50. DOI: 10.1002/dvdy.10083.

Palhière, I., Brochard, M., Moazami-Goudarzi, K., Laloë, D., Amigues, Y., Bed'hom, B., Neuts, É., Leymarie, C., Pantano, T., Cribiu, E. P., Bibé, B., Verrier, É., 2008. Impact of strong selection for the PrP major gene on genetic variability of four French sheep breeds (Open Access publication). *Genetics Selection Evolution*, 40, 6, 663. DOI: 10.1186/1297-9686-40-6-663.

Palomino, D. C. T., Marti, L. C., Palomino, D. C. T., Marti, L. C., 2015. Chemokines and immunity. *Einstein (São Paulo)*, 13, 3, 469-473. DOI: 10.1590/S1679-45082015RB3438.

Pezet, A., Favre, H., Kelly, P. A., Edery, M., 1999. Inhibition and Restoration of Prolactin Signal Transduction by Suppressors of Cytokine Signaling. *Journal of Biological Chemistry*, 274, 35, 24497-24502. DOI: 10.1074/jbc.274.35.24497.

Piacère, A., Bouloc-Duval, N., Sigwald, J. P., Larzul, C., Manfredi, E., 1997. Utilisation de l'index combiné caprin et du polymorphisme de la caséine alpha s1 dans le schéma de sélection caprin. Paris (France).

Piessevaux, J., Lavens, D., Montoye, T., Wauman, J., Catteeuw, D., Vandekerckhove, J., Belsham, D., Peelman, F., Tavernier, J., 2006. Functional Cross-modulation between SOCS Proteins Can Stimulate Cytokine Signaling. *Journal of Biological Chemistry*, 281, 44, 32953-32966. DOI: 10.1074/jbc.M600776200.

Piessevaux, J., Lavens, D., Peelman, F., Tavernier, J., 2008. The many faces of the SOCS box. *Cytokine & Growth Factor Reviews*, 19, 5, 371-381. DOI: 10.1016/j.cytogfr.2008.08.006.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R-core Team, 2019. *nlme: Linear and Nonlinear Mixed Effects Models*.

Posselt, G., Schwarz, H., Duschl, A., Horejs-Hoeck, J., 2011. Suppressor of Cytokine Signaling 2 Is a Feedback Inhibitor of TLR-Induced Activation in Human Monocyte-Derived

Dendritic Cells. *The Journal of Immunology*, 187, 6, 2875-2884. DOI: 10.4049/jimmunol.1003348.

Poulsen, L. la C., Englund, A. L. M., Wissing, M. L. M., Yding Andersen, C., Borup, R., Grøndahl, M. L., 2019. Human granulosa cells function as innate immune cells executing an inflammatory reaction during ovulation: a microarray analysis. *Molecular and Cellular Endocrinology*, 486, 34-46. DOI: 10.1016/j.mce.2019.02.014.

Pritchard, T., Coffey, M., Mrode, R., Wall, E., 2013. Genetic parameters for production, health, fertility and longevity traits in dairy cows. *Animal: An International Journal of Animal Bioscience*, 7, 1, 34-46. DOI: 10.1017/S1751731112001401.

Qing, Y., Costa-Pereira, A. P., Watling, D., Stark, G. R., 2005. Role of Tyrosine 441 of Interferon- γ Receptor Subunit 1 in SOCS-1-mediated Attenuation of STAT1 Activation. *Journal of Biological Chemistry*, 280, 3, 1849-1853. DOI: 10.1074/jbc.M409863200.

Qiu, X., Zheng, J., Guo, X., Gao, X., Liu, H., Tu, Y., Zhang, Y., 2013. Reduced expression of SOCS2 and SOCS6 in hepatocellular carcinoma correlates with aggressive tumor progression and poor prognosis. *Molecular and Cellular Biochemistry*, 378, 1, 99-106. DOI: 10.1007/s11010-013-1599-5.

Quinton, C. D., Hely, F. S., Amer, P. R., Byrne, T. J., Cromie, A. R., 2018. Prediction of effects of beef selection indexes on greenhouse gas emissions. *Animal: An International Journal of Animal Bioscience*, 12, 5, 889-897. DOI: 10.1017/S1751731117002373.

R Development Core Team, 2005. *R: A language and environment for statistical computing*. Disponible sur: <http://www.R-project.org/> (consulté le 01/06/2018).

Ram, P. A., Waxman, D. J., 1999. SOCS/CIS Protein Inhibition of Growth Hormone-stimulated STAT5 Signaling by Multiple Mechanisms. *Journal of Biological Chemistry*, 274, 50, 35553-35561. DOI: 10.1074/jbc.274.50.35553.

2000. Role of the Cytokine-inducible SH2 Protein CIS in Desensitization of STAT5b Signaling by Continuous Growth Hormone. *Journal of Biological Chemistry*, 275, 50, 39487-39496. DOI: 10.1074/jbc.M004755200.

Ransome, M. I., Goldshmit, Y., Bartlett, P. F., Waters, M. J., Turnley, A. M., 2004. Comparative analysis of CNS populations in knockout mice with altered growth hormone responsiveness. *European Journal of Neuroscience*, 19, 8, 2069-2079. DOI: 10.1111/j.0953-816X.2004.03308.x.

Rastislav, M., Mangesh, B., 2012. BoLA-DRB3 exon 2 mutations associated with paratuberculosis in cattle. *The Veterinary Journal*, 192, 3, 517-519. DOI: 10.1016/j.tvjl.2011.07.005.

Rawlings, J. S., Rosler, K. M., Harrison, D. A., 2004. The JAK/STAT signaling pathway. *Journal of Cell Science*, 117, 8, 1281-1283. DOI: 10.1242/jcs.00963.

- Rico-Bautista, E., Flores-Morales, A., Fernández-Pérez, L.,** 2006. Suppressor of cytokine signaling (SOCS) 2, a protein with multiple functions. *Cytokine & Growth Factor Reviews*, 17, 6, 431-439. DOI: 10.1016/j.cytogfr.2006.09.008.
- Robinson, M. D., McCarthy, D. J., Smyth, G. K.,** 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 1, 139-140. DOI: 10.1093/bioinformatics/btp616.
- Robinson, M. D., Oshlack, A.,** 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11, 3, R25. DOI: 10.1186/gb-2010-11-3-r25.
- Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., Servin, B.,** 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics*, 19. DOI: 10.1186/s12864-018-4447-x.
- Rupp, R., Foucras, G.,** 2010. Genetics of mastitis in dairy ruminants. Dans Bishop, S. C., Axford, R. F. E., Nicholas, F. W., Owen, J. B. (dir.). *Breeding for disease resistance in farm animals*. CABI, Wallingford, 183-212. DOI: 10.1079/9781845935559.0183.
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., Portes, D., Woloszyn, F., Bouchez, O., Tabouret, G., Lebastard, M., Caubet, C., Foucras, G., Tosser-Klopp, G.,** 2015. A point mutation in suppressor of cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genetics*, 11, 12, e1005629. DOI: 10.1371/journal.pgen.1005629.
- Salzberg, S. L.,** 2018. Open questions: How many genes do we have? *BMC Biology*, 16, 1, 94. DOI: 10.1186/s12915-018-0564-x.
- Sandra, O., Bataillon, I., Roux, P., Martal, J., Charpigny, G., Renaud, P., Bolifraud, P., Germain, G., Al-Gubory, K. H.,** 2005. Suppressor of cytokine signalling (SOCS) genes are expressed in the endometrium and regulated by conceptus signals during early pregnancy in the ewe. *Journal of Molecular Endocrinology*, 34, 3, 637-644. DOI: 10.1677/jme.1.01667.
- Sartelet, A., Klingbeil, P., Franklin, C. K., Fasquelle, C., Géron, S., Isacke, C. M., Georges, M., Charlier, C.,** 2012. Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection? *Animal Genetics*, 43, 5, 604-607. DOI: 10.1111/j.1365-2052.2011.02311.x.
- Schindler, C., Plumlee, C.,** 2008. Interferons pen the JAK-STAT pathway. *Seminars in cell & developmental biology*, 19, 4, 311-318. DOI: 10.1016/j.semcd.2008.08.010.
- Seki, Y., Hayashi, K., Matsumoto, A., Seki, N., Tsukada, J., Ransom, J., Naka, T., Kishimoto, T., Yoshimura, A., Kubo, M.,** 2002. Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proceedings of the National Academy of Sciences*, 99, 20, 13003-13008. DOI: 10.1073/pnas.202477099.

Selvaggi, M., Laudadio, V., Dario, C., Tufarelli, V., 2014a. Investigating the genetic polymorphism of sheep milk proteins: a useful tool for dairy production. *Journal of the Science of Food and Agriculture*, 94, 15, 3090-3099. DOI: 10.1002/jsfa.6750.

2014b. Major proteins in goat milk: an updated overview on genetic variability. *Molecular Biology Reports*, 41, 2, 1035-1048. DOI: 10.1007/s11033-013-2949-9.

Shuai, K., Liu, B., 2003. Regulation of JAK–STAT signalling in the immune system. *Nature Reviews Immunology*, 3, 11, 900. DOI: 10.1038/nri1226.

Shuster, D. E., Kehrli, M. E., Ackermann, M. R., Gilbert, R. O., 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences*, 89, 19, 9225-9229. DOI: 10.1073/pnas.89.19.9225.

Silva, B. D. M., Castro, E. A., Souza, C. J. H., Paiva, S. R., Sartori, R., Franco, M. M., Azevedo, H. C., Silva, T. a. S. N., Vieira, A. M. C., Neves, J. P., Melo, E. O., 2011. A new polymorphism in the Growth and Differentiation Factor 9 (GDF9) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. *Animal Genetics*, 42, 1, 89-92. DOI: 10.1111/j.1365-2052.2010.02078.x.

Sonstegard, T. S., Cole, J. B., VanRaden, P. M., Tassell, C. P. V., Null, D. J., Schroeder, S. G., Bickhart, D., McClure, M. C., 2013. Identification of a Nonsense Mutation in CWC15 Associated with Decreased Reproductive Efficiency in Jersey Cattle. *PLOS ONE*, 8, 1, e54872. DOI: 10.1371/journal.pone.0054872.

Sordillo, L. M., 2018. Mammary Gland Immunobiology and Resistance to Mastitis. *Veterinary Clinics of North America: Food Animal Practice*, 34, 3, 507-523. DOI: 10.1016/j.cvfa.2018.07.005.

Souza, C. J. H., McNeilly, A. S., Benavides, M. V., Melo, E. O., Moraes, J. C. F., 2014. Mutation in the protease cleavage site of GDF9 increases ovulation rate and litter size in heterozygous ewes and causes infertility in homozygous ewes. *Animal Genetics*, 45, 5, 732-739. DOI: 10.1111/age.12190.

Souza, C. J., MacDougall, C., MacDougall, C., Campbell, B. K., McNeilly, A. S., Baird, D. T., 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *Journal of Endocrinology*, 169, 2, R1-R6. DOI: 10.1677/joe.0.169r001.

Sovio, U., Bennett, A. J., Millwood, I. Y., Molitor, J., O'Reilly, P. F., Timpson, N. J., Kaakinen, M., Laitinen, J., Haukka, J., Pillas, D., Tzoulaki, I., Molitor, J., Hoggart, C., Coin, L. J. M., Whittaker, J., Pouta, A., Hartikainen, A.-L., Freimer, N. B., Widen, E., Peltonen, L., Elliott, P., McCarthy, M. I., Jarvelin, M.-R., 2009. Genetic Determinants of Height Growth Assessed Longitudinally from Infancy to Adulthood in the Northern Finland Birth Cohort 1966. *PLOS Genetics*, 5, 3, e1000409. DOI: 10.1371/journal.pgen.1000409.

Sporri, B., Kovanen, P. E., Sasaki, A., Yoshimura, A., Leonard, W. J., 2001. JAB/SOCS1/SSI-1 is an interleukin-2-induced inhibitor of IL-2 signaling. *Blood*, 97, 1, 221-226.

- Starr, R., Fuchsberger, M., Lau, L. S., Uldrich, A. P., Goradia, A., Willson, T. A., Verhagen, A. M., Alexander, W. S., Smyth, M. J., 2009.** SOCS-1 binding to tyrosine 441 of IFN-gamma receptor subunit 1 contributes to the attenuation of IFN-gamma signaling in vivo. *Journal of Immunology (Baltimore, Md.: 1950)*, 183, 7, 4537-4544. DOI: 10.4049/jimmunol.0901010.
- Starr, R., Willson, T. A., Viney, E. M., Murray, L. J., Rayner, J. R., Jenkins, B. J., Gonda, T. J., Alexander, W. S., Metcalf, D., Nicola, N. A., Hilton, D. J., 1997.** A family of cytokine-inducible inhibitors of signalling. *Nature*, 387, 6636, 917-921. DOI: 10.1038/43206.
- Suárez-Vega, A., Gutiérrez-Gil, B., Benavides, J., Perez, V., Tosser-Klopp, G., Klopp, C., Keenel, S. J., Arranz, J. J., 2015.** Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PLoS One*, 10, 5, e0126416. DOI: 10.1371/journal.pone.0126416.
- Sugimoto, M., Fujikawa, A., Womack, J. E., Sugimoto, Y., 2006.** Evidence that bovine forebrain embryonic zinc finger-like gene influences immune response associated with mastitis resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 17, 6454-6459. DOI: 10.1073/pnas.0601015103.
- Sutherland, J. M., Keightley, R. A., Nixon, B., Roman, S. D., Robker, R. L., Russell, D. L., McLaughlin, E. A., 2012.** Suppressor of cytokine signaling 4 (SOCS4): Moderator of ovarian primordial follicle activation. *Journal of Cellular Physiology*, 227, 3, 1188-1198. DOI: 10.1002/jcp.22837.
- Sutherland, K. D., Lindeman, G. J., Choong, D. Y. H., Wittlin, S., Brentzell, L., Phillips, W., Campbell, I. G., Visvader, J. E., 2004.** Differential hypermethylation of SOCS genes in ovarian and breast carcinomas. *Oncogene*, 23, 46, 7726. DOI: 10.1038/sj.onc.1207787.
- Takahashi, Y., Carpino, N., Cross, J. C., Torres, M., Parganas, E., Ihle, J. N., 2003.** SOCS3: an essential regulator of LIF receptor signaling in trophoblast giant cell differentiation. *The EMBO Journal*, 22, 3, 372-384. DOI: 10.1093/emboj/cdg057.
- Tannahill, G. M., Elliott, J., Barry, A. C., Hibbert, L., Cacalano, N. A., Johnston, J. A., 2005.** SOCS2 Can Enhance Interleukin-2 (IL-2) and IL-3 Signaling by Accelerating SOCS3 Degradation. *Molecular and Cellular Biology*, 25, 20, 9115-9126. DOI: 10.1128/MCB.25.20.9115-9126.2005.
- Teissier, M., Larroque, H., Robert-Granié, C., 2018.** Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. *Genetics Selection Evolution*, 50, 1. DOI: 10.1186/s12711-018-0400-3.
- Teissier, M., Larroque, H., Robert-Granié, C., 2019.** Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *Journal of Dairy Science*, 102, 4, 3142-3154. DOI: 10.3168/jds.2018-15650.
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A. H., Holm, L.-E., Nielsen, V. H., Agerholm, J. S., Arnbjerg, J., Bendixen, C., 2006.** A missense mutation in the bovine

SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Research*, 16, 1, 97-105. DOI: 10.1101/gr.3690506.

Tortereau, F., 2016. Conséquences de la sélection pour la résistance à la tremblante sur la diversité génétique du génome ovin. *Conséquences de la sélection pour la résistance à la tremblante sur la diversité génétique du génome ovin.* . Paris.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28, 5, 511-515. DOI: 10.1038/nbt.1621.

Turnley, A. M., Faux, C. H., Rietze, R. L., Coonan, J. R., Bartlett, P. F., 2002. Suppressor of cytokine signaling 2 regulates neuronal differentiation by inhibiting growth hormone signaling. *Nature Neuroscience*, 5, 11, 1155. DOI: 10.1038/nn954.

Udy, G. B., Towers, R. P., Snell, R. G., Wilkins, R. J., Park, S.-H., Ram, P. A., Waxman, D. J., Davey, H. W., 1997. Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression. *Proceedings of the National Academy of Sciences*, 94, 14, 7239-7244. DOI: 10.1073/pnas.94.14.7239.

Usman, T., Yu, Y., Liu, C., Wang, X., Zhang, Q., Wang, Y., 2014. Genetic effects of single nucleotide polymorphisms in JAK2 and STAT5A genes on susceptibility of Chinese Holsteins to mastitis. *Molecular Biology Reports*, 41, 12, 8293-8301. DOI: 10.1007/s11033-014-3730-4.

Våge, D. I., Boman, I. A., 2010. A nonsense mutation in the beta-carotene oxygenase 2 (BCO2) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (*Ovis aries*). *BMC genetics*, 11, 10. DOI: 10.1186/1471-2156-11-10.

Våge, D. I., Husdal, M., Kent, M. P., Klemetsdal, G., Boman, I. A., 2013. A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. *BMC genetics*, 14, 1. DOI: 10.1186/1471-2156-14-1.

Vesterlund, M., Zadjali, F., Persson, T., Nielsen, M. L., Kessler, B. M., Norstedt, G., Flores-Morales, A., 2011. The SOCS2 Ubiquitin Ligase Complex Regulates Growth Hormone Receptor Levels. *PLOS ONE*, 6, 9, e25358. DOI: 10.1371/journal.pone.0025358.

Vignais, M. L., 1997. Protéines JAK et STAT dans la signalisation cellulaire. *Médecine/Science*, 13, 1277-1284. DOI: 10.4267/10608/546.

Vignal, A., 2011. Etat actuel du séquençage et de la connaissance du génome des espèces animales. *INRA Productions Animales*, 24, 4, 387-404.

Villarino, A. V., Kanno, Y., O'Shea, J. J., 2017. Mechanisms and consequences of Jak-STAT signaling in the immune system. *Nature Immunology*, 18, 4, 374-384. DOI: 10.1038/ni.3691.

Visscher, P. M., Goddard, M. E., 2019. From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics*, 211, 4, 1125-1130. DOI: 10.1534/genetics.118.301594.

Wakioka, T., Sasaki, A., Mitsui, K., Yokouchi, M., Inoue, A., Komiya, S., Yoshimura, A., 1999. APS, an adaptor protein containing Pleckstrin homology (PH) and Src homology-2 (SH2)

domains inhibits the JAK-STAT pathway in collaboration with c-Cbl. *Leukemia*, 13, 5, 760-767.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Muir, W. M., 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, 94, 2, 73-83. DOI: 10.1017/S0016672312000274.

Wieczorek, M., Ginter, T., Brand, P., Heinzl, T., Krämer, O. H., 2012. Acetylation modulates the STAT signaling code. *Cytokine & Growth Factor Reviews*, 23, 6, 293-305. DOI: 10.1016/j.cytogfr.2012.06.005.

Wikman, H., Kettunen, E., Seppänen, J. K., Karjalainen, A., Hollmén, J., Anttila, S., Knuutila, S., 2002. Identification of differentially expressed genes in pulmonary adenocarcinoma by using cDNA array. *Oncogene*, 21, 37, 5804. DOI: 10.1038/sj.onc.1205726.

Wilks, A. F., Harpur, A. G., Kurban, R. R., Ralph, S. J., Zürcher, G., Ziemiecki, A., 1991. Two novel protein-tyrosine kinases, each with a second phosphotransferase-related catalytic domain, define a new class of protein kinase. *Molecular and Cellular Biology*, 11, 4, 2057-2065.

Wilson, T., Wu, X.-Y., Juengel, J. L., Ross, I. K., Lumsden, J. M., Lord, E. A., Dodds, K. G., Walling, G. A., McEwan, J. C., O'Connell, A. R., McNatty, K. P., Montgomery, G. W., 2001. Highly Prolific Booroola Sheep Have a Mutation in the Intracellular Kinase Domain of Bone Morphogenetic Protein IB Receptor (ALK-6) That Is Expressed in Both Oocytes and Granulosa Cells. *Biology of Reproduction*, 64, 4, 1225-1235. DOI: 10.1095/biolreprod64.4.1225.

Winter, A., Krämer, W., Werner, F. A. O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J. E., Thaller, G., Fries, R., 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proceedings of the National Academy of Sciences*, 99, 14, 9300-9305. DOI: 10.1073/pnas.142293799.

Wójcik, M., Krawczyńska, A., Antushevich, H., Herman, A. P., 2018. Post-Receptor Inhibitors of the GHR-JAK2-STAT Pathway in the Growth Hormone Signal Transduction. *International Journal of Molecular Sciences*, 19, 7. DOI: 10.3390/ijms19071843.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M. L., Croteau-Chonka, D. C., Day, F. R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A. U., Karjalainen, J., Lo, K. S., Locke, A. E., Mägi, R., Mihailov, E., Porcu, E., Randall, J. C., Scherag, A., Vinkhuyzen, A. A. E., Westra, H.-J., Winkler, T. W., Workalemahu, T., Zhao, J. H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G. B., Feenstra, B., Feitosa, M. F., Fischer, K., Fraser, R. M., Goel, A., Gong, J., Justice, A. E., Kanoni, S., Kleber, M. E., Kristiansson, K., Lim, U., Lotay, V., Lui, J. C., Mangino, M., Leach, I. M., Medina-Gomez, C., Nalls, M. A., Nyholt, D. R., Palmer, C. D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J. S., Ripke, S., Shungin, D., Stancáková, A., Strawbridge, R. J., Sung, Y. J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S. W., van Setten, J., Van Vliet-Ostaptchouk, J. V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Ärnlöv, J., Arscott, G. M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A. J., Berne, C., Blüher, M., Bolton, J. L., Böttcher, Y., Boyd, H. A., Bruinenberg, M., Buckley, B. M.,

Buyske, S., Caspersen, I. H., Chines, P. S., Clarke, R., Claudi-Boehm, S., Cooper, M., Daw, E. W., De Jong, P. A., Deelen, J., Delgado, G., Denny, J. C., Dhonukshe-Rutten, R., Dimitriou, M., Doney, A. S. F., Dörr, M., Eklund, N., Eury, E., Folkersen, L., Garcia, M. E., Geller, F., Giedraitis, V., Go, A. S., Grallert, H., Grammer, T. B., Gräßler, J., Grönberg, H., de Groot, L. C. P. G. M., Groves, C. J., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hannemann, A., Hartman, C. A., Hassinen, M., Hayward, C., Heard-Costa, N. L., Helmer, Q., Hemani, G., Henders, A. K., Hillege, H. L., Hlatky, M. A., Hoffmann, W., Hoffmann, P., Holmen, O., Houwing-Duistermaat, J. J., Illig, T., Isaacs, A., James, A. L., Jeff, J., Johansen, B., Johansson, Å., Jolley, J., Juliusdottir, T., Junttila, J., Kho, A. N., Kinnunen, L., Klopp, N., Kocher, T., Kratzer, W., Lichtner, P., Lind, L., Lindström, J., Lobbens, S., Lorentzon, M., Lu, Y., Lyssenko, V., Magnusson, P. K. E., Mahajan, A., Maillard, M., McArdle, W. L., McKenzie, C. A., McLachlan, S., McLaren, P. J., Menni, C., Merger, S., Milani, L., Moayyeri, A., Monda, K. L., Morken, M. A., Müller, G., Müller-Nurasyid, M., Musk, A. W., Narisu, N., Nauck, M., Nolte, I. M., Nöthen, M. M., Oozageer, L., Pilz, S., Rayner, N. W., Renstrom, F., Robertson, N. R., Rose, L. M., Roussel, R., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F. R., Schunkert, H., Scott, R. A., Sehmi, J., Seufferlein, T., Shi, J., Silventoinen, K., Smit, J. H., Smith, A. V., Smolonska, J., Stanton, A. V., Stirrups, K., Stott, D. J., Stringham, H. M., Sundström, J., Swertz, M. A., Syvänen, A.-C., Tayo, B. O., Thorleifsson, G., Tyrer, J. P., van Dijk, S., van Schoor, N. M., van der Velde, N., van Heemst, D., van Oort, F. V. A., Vermeulen, S. H., Verweij, N., Vonk, J. M., Waite, L. L., Waldenberger, M., Wennauer, R., Wilkens, L. R., Willenborg, C., Wilsgaard, T., Wojczynski, M. K., Wong, A., Wright, A. F., Zhang, Q., Arveiler, D., Bakker, S. J. L., Beilby, J., Bergman, R. N., Bergmann, S., Biffar, R., Blangero, J., Boomsma, D. I., Bornstein, S. R., Bovet, P., Brambilla, P., Brown, M. J., Campbell, H., Caulfield, M. J., Chakravarti, A., Collins, R., Collins, F. S., Crawford, D. C., Cupples, L. A., Danesh, J., de Faire, U., den Ruijter, H. M., Erbel, R., Erdmann, J., Eriksson, J. G., Farrall, M., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N. G., Forrester, T., Gansevoort, R. T., Gejman, P. V., Gieger, C., Golay, A., Gottesman, O., Gudnason, V., Gyllenstein, U., Haas, D. W., Hall, A. S., Harris, T. B., Hattersley, A. T., Heath, A. C., Hengstenberg, C., Hicks, A. A., Hindorff, L. A., Hingorani, A. D., Hofman, A., Hovingh, G. K., Humphries, S. E., Hunt, S. C., Hyppönen, E., Jacobs, K. B., Jarvelin, M.-R., Jousilahti, P., Jula, A. M., Kaprio, J., Kastelein, J. J. P., Kayser, M., Kee, F., Keinanen-Kiukaanniemi, S. M., Kiemeny, L. A., Kooner, J. S., Kooperberg, C., Koskinen, S., Kovacs, P., Kraja, A. T., Kumari, M., Kuusisto, J., Lakka, T. A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lupoli, S., Madden, P. A. F., Männistö, S., Manunta, P., Marette, A., Matisse, T. C., McKnight, B., Meitinger, T., Moll, F. L., Montgomery, G. W., Morris, A. D., Morris, A. P., Murray, J. C., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Ouwehand, W. H., Pasterkamp, G., Peters, A., Pramstaller, P. P., Price, J. F., Qi, L., Raitakari, O. T., Rankinen, T., Rao, D. C., Rice, T. K., Ritchie, M., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schwarz, P. E. H., Sebert, S., Sever, P., Shuldiner, A. R., Sinisalo, J., Steinthorsdottir, V., Stolk, R. P., Tardif, J.-C., Tönjes, A., Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.-C., The Electronic Medical Records and Genomics (eMERGE) Consortium, The MIGen Consortium, The PAGE Consortium, The LifeLines Cohort Study, Amouyel, P., Asselbergs, F. W., Assimes, T. L., Bochud, M., Boehm, B. O., Boerwinkle, E., Bottinger, E. P., Bouchard, C., Cauchi, S., Chambers, J. C., Chanock, S. J., Cooper, R. S., de Bakker, P. I. W., Dedoussis, G., Ferrucci, L., Franks, P. W., Froguel, P., Groop, L. C., Haiman, C. A., Hamsten, A., Hayes, M. G., Hui, J., Hunter, D. J., Hveem, K., Jukema, J. W., Kaplan, R. C., Kivimäki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N. G., März, W., Melbye, M., Moebus, S., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N. A., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Powell, J. E.,

Power, C., Quertermous, T., Rauramaa, R., Reinmaa, E., Ridker, P. M., Rivadeneira, F., Rotter, J. I., Saaristo, T. E., Saleheen, D., Schlessinger, D., Slagboom, P. E., Snieder, H., Spector, T. D., Strauch, K., Stumvoll, M., Tuomilehto, J., Uusitupa, M., van der Harst, P., Völzke, H., Walker, M., Wareham, N. J., Watkins, H., Wichmann, H.-E., Wilson, J. F., Zanen, P., Deloukas, P., Heid, I. M., Lindgren, C. M., Mohlke, K. L., Speliotes, E. K., Thorsteinsdottir, U., Barroso, I., Fox, C. S., North, K. E., Strachan, D. P., Beckmann, J. S., Berndt, S. I., Boehnke, M., Borecki, I. B., McCarthy, M. I., Metspalu, A., Stefansson, K., Uitterlinden, A. G., van Duijn, C. M., Franke, L., Willer, C. J., Price, A. L., Lettre, G., Loos, R. J. F., Weedon, M. N., Ingelsson, E., O'Connell, J. R., Abecasis, G. R., Chasman, D. I., Goddard, M. E., Visscher, P. M., Hirschhorn, J. N., Frayling, T. M., 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46, 11, 1173-1186. DOI: 10.1038/ng.3097.

Wormald, S., Hilton, D. J., 2004. Inhibitors of Cytokine Signal Transduction. *Journal of Biological Chemistry*, 279, 2, 821-824. DOI: 10.1074/jbc.R300030200.

Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., MacArthur, D. G., Quail, M. A., Carter, N. P., Yang, H., Tyler-Smith, C., 2009. Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. *Current Biology*, 19, 17, 1453-1457. DOI: 10.1016/j.cub.2009.07.032.

Yang, J., Liu, X., Wang, D., Ning, C., Wang, H., Zhang, Q., Jiang, L., 2017. Functional validation of GPIHBP1 and identification of a functional mutation in GPIHBP1 for milk fat traits in dairy cattle. *Scientific Reports*, 7, 1, 8546. DOI: 10.1038/s41598-017-08668-6.

Yasukawa, H., Misawa, H., Sakamoto, H., Masuhara, M., Sasaki, A., Wakioka, T., Ohtsuka, S., Imaizumi, T., Matsuda, T., Ihle, J. N., Yoshimura, A., 1999. The JAK-binding protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop. *The EMBO Journal*, 18, 5, 1309-1320. DOI: 10.1093/emboj/18.5.1309.

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., 2018. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics*, 27, 20, 3641-3649. DOI: 10.1093/hmg/ddy271.

Yoshimura, A., Ito, M., Chikuma, S., Akanuma, T., Nakatsukasa, H., 2018. Negative Regulation of Cytokine Signaling in Immunity. *Cold Spring Harbor Perspectives in Biology*, 10, 7, a028571. DOI: 10.1101/cshperspect.a028571.

Yoshimura, A., Ohkubo, T., Kiguchi, T., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Hara, T., Miyajima, A., 1995. A novel cytokine-inducible gene CIS encodes an SH2-containing protein that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin receptors. *The EMBO Journal*, 14, 12, 2816-2826. DOI: 10.1002/j.1460-2075.1995.tb07281.x.

Zeder, M. A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 33, 11597-11604. DOI: 10.1073/pnas.0801317105.

Zhang, J.-G., Farley, A., Nicholson, S. E., Willson, T. A., Zugaro, L. M., Simpson, R. J., Moritz, R. L., Cary, D., Richardson, R., Hausmann, G., Kile, B. T., Kent, S. B. H., Alexander, W. S., Metcalf, D., Hilton, D. J., Nicola, N. A., Baca, M., 1999. The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 5, 2071-2076.

Zhang, Q., Chen, J. Q., Lin, J., Yu, Q. H., Yu, H. Q., Xu, X. J., Liu, G. H., Yang, Q., 2014. Production GH transgenic goat improving mammogenesis by somatic cell nuclear transfer. *Molecular Biology Reports*, 41, 7, 4759-4768. DOI: 10.1007/s11033-014-3347-7.

Zhang, X., Lourenco, D., Aguilar, I., Legarra, A., Misztal, I., 2016. Weighting strategies for single-step genomic BLUP: An iterative approach for accurate calculation of GEBV and GWAS. *Frontiers in Genetics*, 7. DOI: 10.3389/fgene.2016.00151.

Zhou, S., Cai, B., He, C., Wang, Y., Ding, Q., Liu, J., Liu, Y., Ding, Y., Zhao, X., Li, G., Li, C., Yu, H., Kou, Q., Niu, W., Petersen, B., Sonstegard, T., Ma, B., Chen, Y., Wang, X., 2019. Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations. *Frontiers in Genetics*, 10. DOI: 10.3389/fgene.2019.00215.

Zhu, J., Dai, Q., Han, Z., He, H., Mo, R., Chen, G., Chen, Y., Wu, Y., Yang, S., Jiang, F., Chen, W., Sun, Z., Zhong, W., 2013. Expression of SOCSs in human prostate cancer and their association in prognosis. *Molecular and Cellular Biochemistry*, 381, 1, 51-59. DOI: 10.1007/s11010-013-1687-6.

Zuliani, A., Mair, M., Kraševc, M., Lora, I., Brscic, M., Cozzi, G., Leeb, C., Zupan, M., Winckler, C., Bovolenta, S., 2018. A survey of selected animal-based measures of dairy cattle welfare in the Eastern Alps: Toward context-based thresholds. *Journal of Dairy Science*, 101, 2, 1428-1436. DOI: 10.3168/jds.2017-13257.

Productions scientifiques

I. Publications

Oget, C., Allain, C., Portes, D., Foucras, G., Stella, A., Astruc, J.-M., Sarry, J., Tosser-Klopp, G., Rupp, R., 2019. A validation study of loci associated with mastitis resistance in two French dairy sheep breeds. *Genetics Selection Evolution*, 51, 1, 5. DOI: [10.1186/s12711-019-0448-8](https://doi.org/10.1186/s12711-019-0448-8).

Oget, C., Tosser-Klopp, G., Rupp, R., 2019. Genetic and genomic studies in ovine mastitis. *Small Ruminant Research*, 176, 55-64. DOI: [10.1016/j.smallrumres.2019.05.011](https://doi.org/10.1016/j.smallrumres.2019.05.011).

Oget, C., Teissier, M., Astruc, J.-M., Tosser-Klopp, G., Rupp, R., 2019. Alternative methods improve the accuracy of genomic prediction using information from a causal point mutation in a dairy sheep model. *BMC Genomics*, 20, 1, 719. DOI: [10.1186/s12864-019-6068-4](https://doi.org/10.1186/s12864-019-6068-4).

II. Communications orales

Oget, C., Rupp, R., Tosser-Klopp, G., 2018. Pleiotropic effects of the R96C mutation in the dairy sheep *SOCS2* gene. *21th PhD Seminar of INRA Animal Genetics Division*. Mureaux (France).

Oget, C., Teissier, M., Astruc, J.-M., Tosser-Klopp, G., Rupp, R., 2019. Alternative methods improve the accuracy of genomic prediction using the information of a pleiotropic point mutation in a dairy sheep model. *PAG XXVII - Plant and Animal Genome Conference*. San Diego (USA). (Annexe 20)

Oget, C., Rupp, R., Tosser-Klopp, G., 2019. Pleiotropic effect of the R96C mutation in the dairy sheep *SOCS2* gene. *Seminar of GIGA Animal Division*. Liège (Belgique).

III. Posters

Oget, C., Rupp, R., Tosser-Klopp, G., 2017. Pleiotropic effects of the R96C mutation in the dairy sheep *SOCS2* gene. *20th PhD Seminar of INRA Animal Genetics Division*. Rennes (France). (Annexe 21)

Oget, C., Allain, C., Portes, D., Foucras, G., Tosser-Klopp, G., Rupp, R., 2018. Fine mapping of 7 QTL regions in dairy sheep confirms pleiotropic effect of the R96C mutation in the *SOCS2* gene on SCC, bacterial infection, size and milk production. *WCGALP 2018 - World Congress on Genetics Applied to Livestock Production*. Auckland (New Zealand). (Annexe 22)

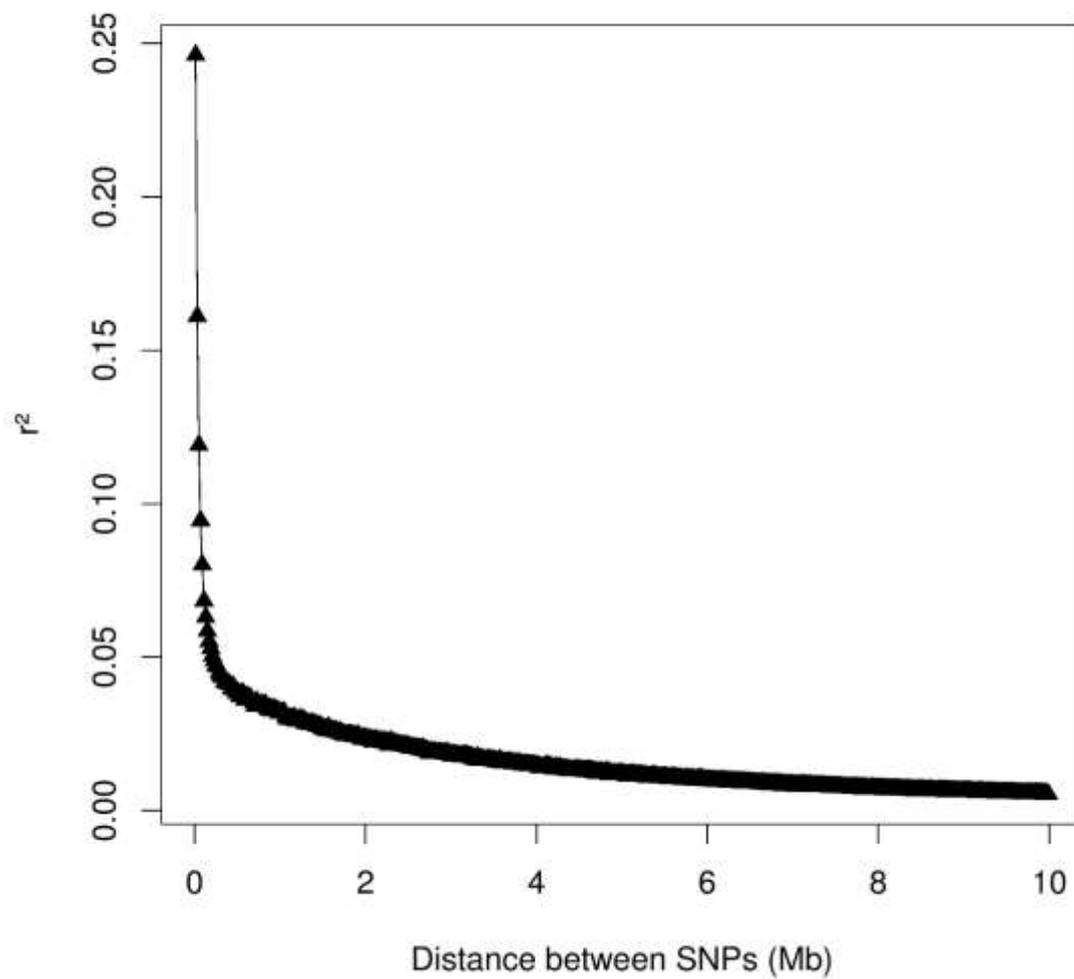
Annexes

Annexe 1 Tableaux additionnels (Table S1) de l'article de synthèse bibliographique ([Chapitre I](#)). Ce tableau présente les calculs effectués pour obtenir une moyenne pondérée de caractères de résistance aux mammites à partir des paramètres génétiques estimés dans la littérature (héritabilités pour le premier tableau et corrélations génétiques avec le caractère de production laitière pour le second)

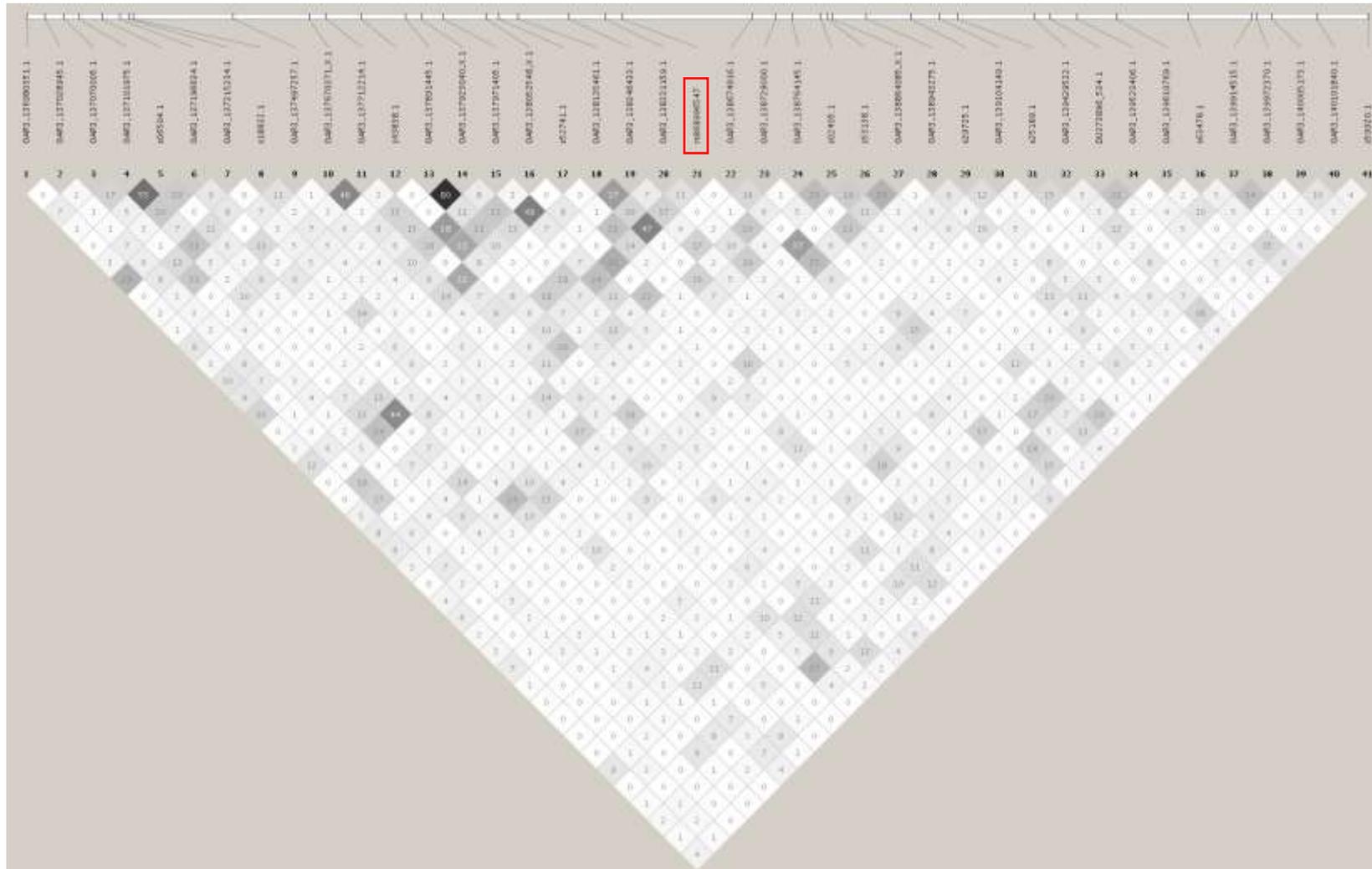
Sheep breed	Most recent reference	Number of ewes	Trait	Heritability	SE	h ² used for the weighted average calculation	SE used for the weighted average calculation	weight = 1/SE	h ² *weight	Number of breeds used for the weighted average calculation	Mean of the given range was chosen
Texel	McLaren et al., 2018	2,957	Av SCS mid-lactation	0.11	0.04	0.110	0.040	25.000	2.750	1	
Lacaune	Allain et al., 2018	377,945	SCS (L1)	0.22	0.004	0.220	0.004	250.000	55.000	1	
Chios	Banos et al., 2017	609	SCS	0.11	0.05	0.110	0.050	20.000	2.200	1	
Valle del Belice	Tolone et al., 2013	2040	SCS	0.09	0.02	0.090	0.020	50.000	4.500	1	
Churra	De la Fuente et al., 2011	3,977	SCS	0.09	0.02	0.090	0.020	50.000	4.500	1	
Sarda, Sarda x Lacaune	Casu et al., 2010	1,587	SCS (L1)	0.26-0.28	0.05	0.270	0.050	20.000	5.400	1	
Manech Red Face	Barillet et al., 2009	58,378	SCS	0.1	0.01-0.02	0.100	0.015	66.667	6.667	1	
Latxa	Legarra and Ugarte, 2005	6,165	SCS	0.13	0.02	0.130	0.020	50.000	6.500	1	
East Friesian	Hamann et al., 2004	1,108	SCS	0.163	0.02	0.163	0.020	50.000	8.150	1	
Manchega	Serrano et al., 2003	10,962	SCS (L1)	0.12	0.01-0.09	0.120	0.050	20.000	2.400	1	Weighted average
TOTAL SCS								601.667	98.067	10	0.163
Texel	McLaren et al., 2018	2,957	Sum CMT mid-lactation	0.09	0.04	0.090	0.040	25.000	2.250	1	
Beldare, Charollais, Suffolk, Texel, Vendéen	O'Brien et al., 2017	3,378	CMT	0.04	0.03	0.040	0.030	33.333	1.333	1	
Chios	Banos et al., 2017	609	CMT	0.12	0.06	0.120	0.060	16.667	2.000	1	Weighted average
TOTAL CMT								75.000	5.583	3	0.074
Lacaune	Allain et al., 2018	518	StaphC	0.18	0.07	0.180	0.070	14.286	2.571	1	
Valle del Belice	Tolone et al., 2016	2,350	InfS by cnS	0.02	0.01	0.020	0.010	100.000	2.000	1	
TOTAL BACTERIOLOGICAL ANALYSES								114.286	4.571	2	0.040
Chios	Banos et al., 2017	609	TBC	0.09	0.05	0.090	0.050	20.000	1.800	1	Weighted average
TOTAL TBC								20.000	1.800	1	0.090
Chios	Banos et al., 2017	609	Dev c/m	0.18	0.1	0.180	0.100	10.000	1.800	1	Moyenne pondérée
TOTAL DEV C/M								10.000	1.800	1	0.180

Sheep breed	Most recent reference	Number of ewes	Trait	Genetic correlation with milk yield	SE	Genetic correlation used for the weighted average calculation with milk yield	SE used for the weighted average calculation	weight = 1/SE	h ² *weight	Number of breeds used for the weighted average calculation	Mean of the given range was chosen
Chios	Banos et al., 2017	609	SCS	-0.12	0.14	-0.120	0.140	7.143	-0.857	1	
Lacaune	Allain et al., 2018	377,945	SCS L1	0.12	0.01	0.120	0.010	100.000	12.000	1	
Manech Red Face	Barillet et al., 2009	58,378	SCS	0.21	0.01-0.03	0.210	0.020	50.000	10.500	1	
Valle del Belice	Riggio et al., 2007	2,277	SCS (L1)	0.23	0.16	0.230	0.160	6.250	1.438	1	
Latxa	Legarra and Ugarte, 2005	6,165	SCS	-0.3	0.07	-0.300	0.070	14.286	-4.286	1	
East Friesian	Hamann et al., 2004	1,108	SCS	0.015	0.02-0.12	0.015	0.070	14.286	0.214	1	
Manchega	Serrano et al., 2003	10,962	SCS (L1)	-0.12	0.01-0.09	-0.120	0.050	20.000	-2.400	1	Weighted average
							TOTAL SCS	211.964	16.609	7	0.078
Chios	Banos et al., 2017	609	CMT	-0.12	0.13	-0.120	0.130	7.692	-0.923	1	Weighted average
							TOTAL CMT	7.692	-0.923	1	-0.120
Valle del Belice	Tolone et al., 2013	2,350	InfS	0.59	0.31	0.590	0.310	3.226	1.903	1	
Lacaune	Allain et al., 2018	518	StaphC	0.31	0.17	0.310	0.170	5.882	1.824	1	Weighted average
							TOTAL BACTERIOLOGICAL ANALYSES	9.108	3.727	2	0.409
Chios	Banos et al., 2017	609	TBC	-0.11	0.14	-0.110	0.140	7.143	-0.786	1	Weighted average
							TOTAL TBC	7.143	-0.786	1	-0.110
Chios	Banos et al., 2017	609	Dev c/m	-0.09	0.19	-0.090	0.190	5.263	-0.474	1	Weighted average
							TOTAL DEV C/M	5.263	-0.474	1	-0.090

Annexe 2 Figure additionnelle (Figure S1) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique ([Chapitre II](#)) : « Visualization of linkage disequilibrium measured as squared correlation coefficient (r^2) according to distance between markers on the 50K ovine SNP chip. »



Annexe 3 Figure additionnelle (Figure S2) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique ([Chapitre II](#)): « Visualization of linkage disequilibrium ($r^2 \times 100$) between the 40 markers closest to the SOCS2 point mutation (rs868996547, in the middle). »

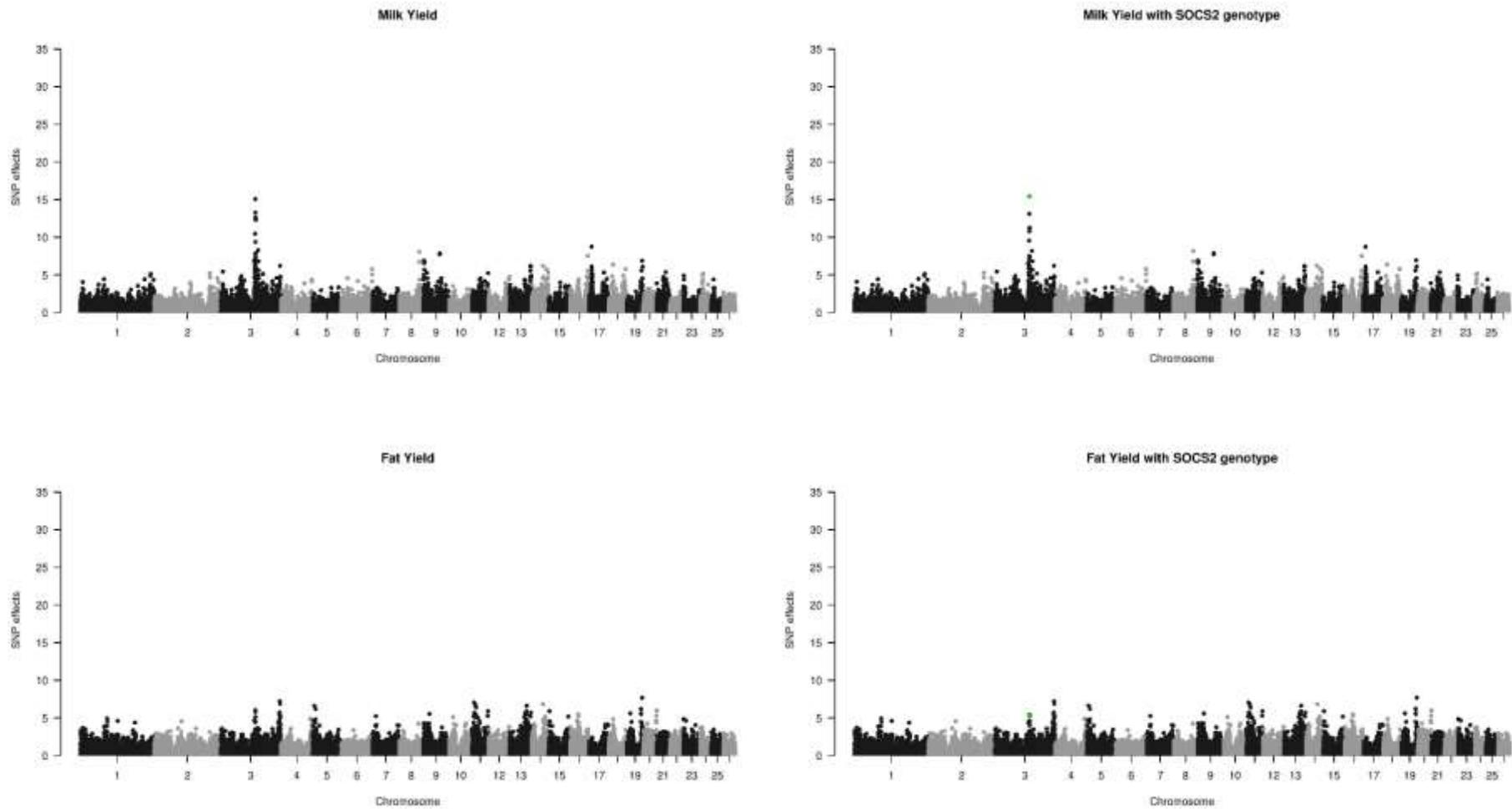


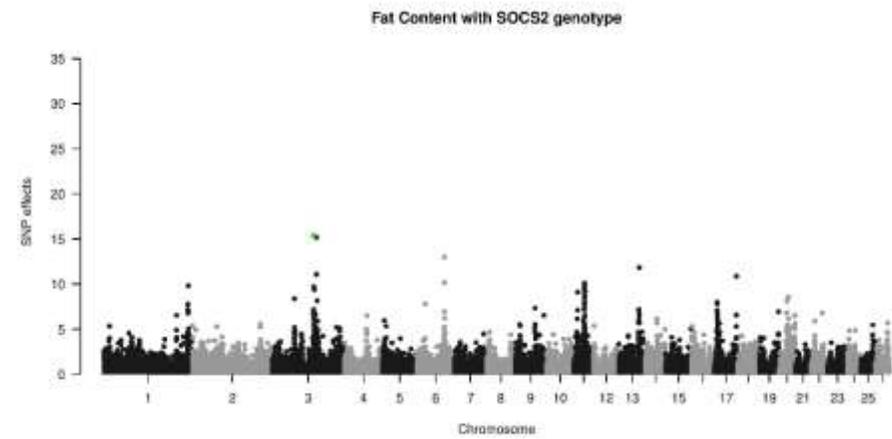
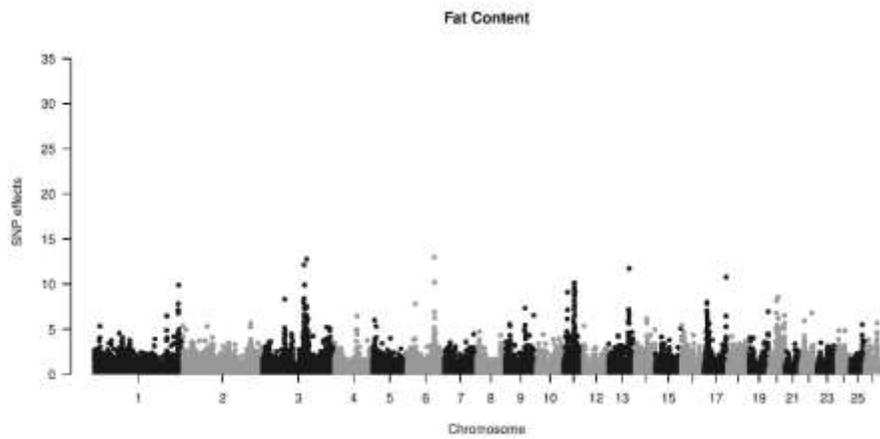
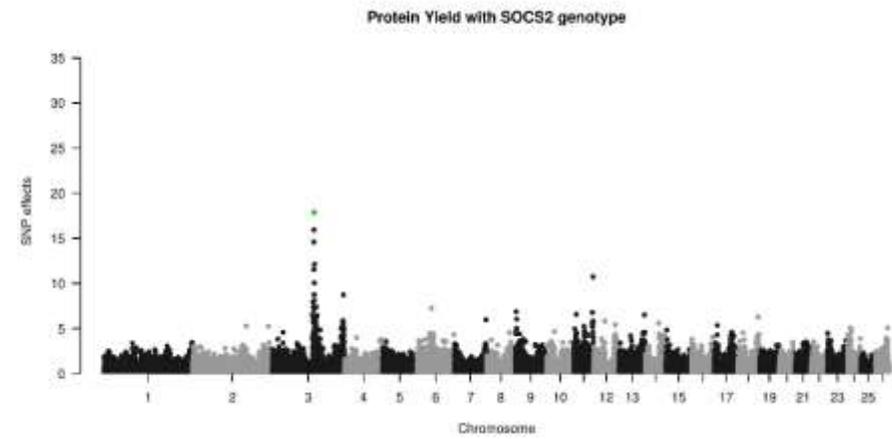
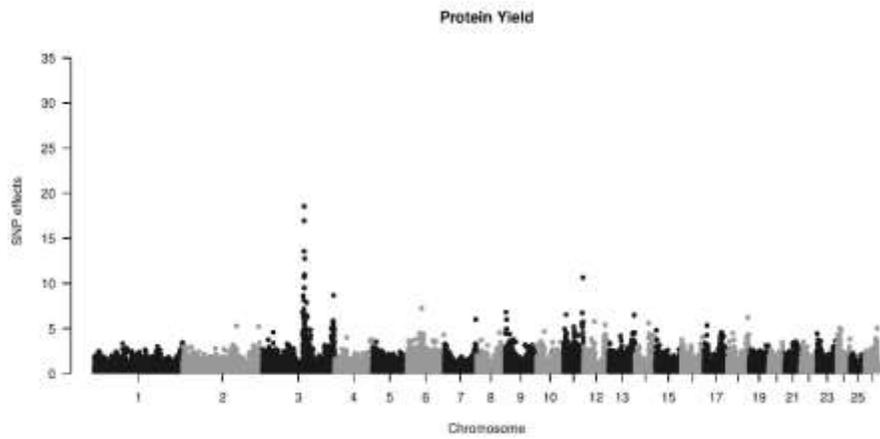
Annexe 4 Figure additionnelle (Figure S3) de l'article scientifique sur la prise en compte de la mutation *SOCS2* dans le contexte de la sélection génomique ([Chapitre II](#)) : « Components estimations according to the different models. One-trait methods correspond to equations (1) and (2) and two-traits methods to equations (3) and (4). »

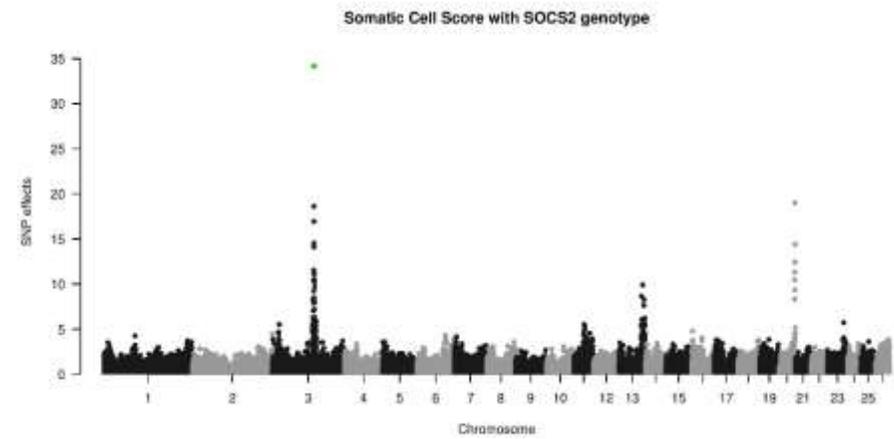
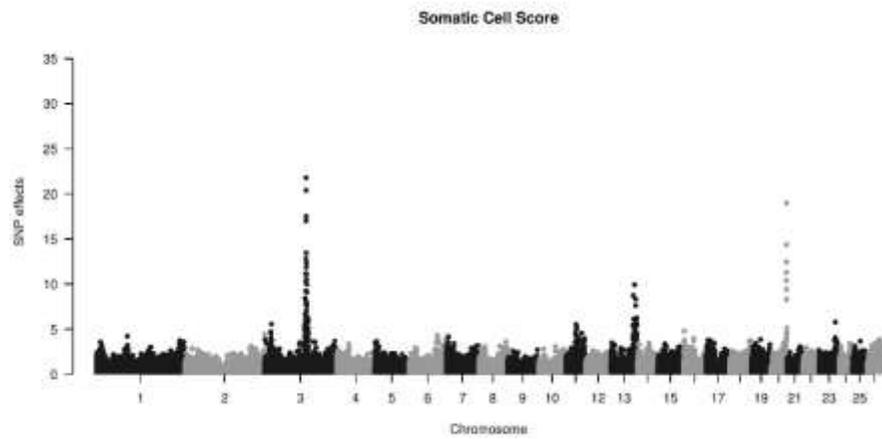
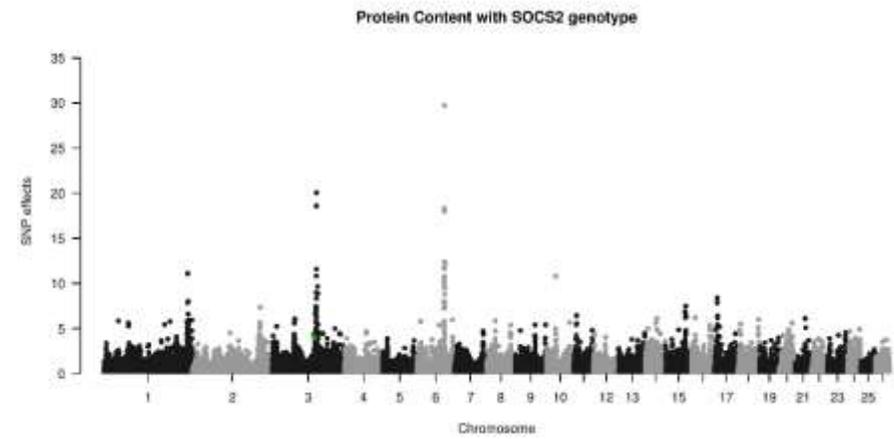
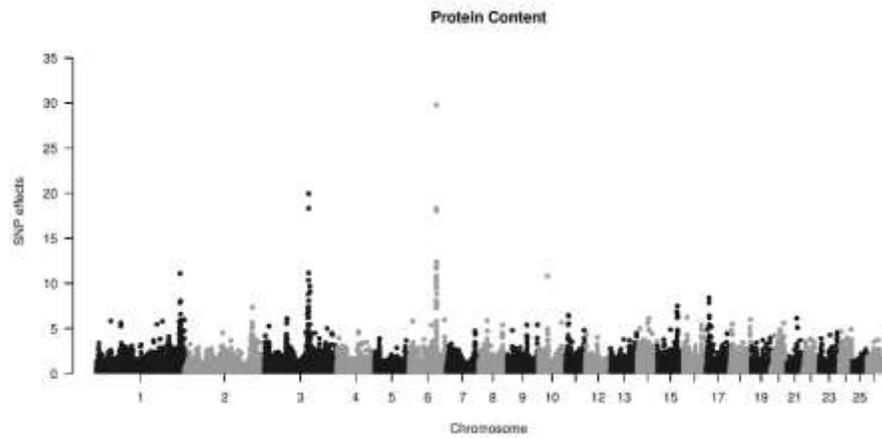
Method	Component	MY	FY	PY	FC	PC	SCS	TA	UC	UD									
One-trait pedigree-based	$\hat{\sigma}_g^2$	1425.00 ± 1.07	604.17 ± 0.45	327.80 ± 0.24	25.86 ± 0.14	8.63 ± 3.98E-02	0.38 ± 4.78E-03	0.37 ± 5.25E-03	0.45 ± 7.08E-03	0.10 ± 1.61E-03									
	$\hat{\sigma}_p^2$	977.98 ± 0.65	311.72 ± 0.29	168.43 ± 0.15	4.96 ± 9.41E-02	1.30 ± 2.37E-02	0.33 ± 3.87E-03												
	$\hat{\sigma}_\varepsilon^2$	1432.00 ± 0.15	716.24 ± 0.12	342.93 ± 5.63E-02	21.26 ± 3.64E-02	4.13 ± 6.93E-03	1.46 ± 2.69E-03	0.58 ± 3.43E-03	0.87 ± 5.15E-03	0.25 ± 1.28E-03									
	h^2	0.37	0.37	0.39	0.50	0.61	0.17	0.39	0.34	0.27									
One-trait genomic-based	$\hat{\sigma}_g^2$	1431.60 ± 1.07	607.51 ± 0.45	329.84 ± 0.24	25.92 ± 0.14	8.66 ± 3.94E-02	0.38 ± 4.73E-03	0.37 ± 5.38E-03	0.45 ± 7.30E-03	0.10 ± 1.67E-03									
	$\hat{\sigma}_p^2$	975.35 ± 0.65	310.55 ± 0.29	167.68 ± 0.15	4.94 ± 9.00E-02	1.29 ± 2.28E-02	0.33 ± 3.75E-03												
	$\hat{\sigma}_\varepsilon^2$	1432.00 ± 0.15	716.24 ± 0.12	342.93 ± 5.66E-02	21.26 ± 3.65E-02	4.13 ± 6.93E-03	1.46 ± 2.71E-03	0.58 ± 3.47E-03	0.87 ± 5.23E-03	0.25 ± 1.30E-03									
	h^2	0.37	0.37	0.39	0.50	0.61	0.18	0.39	0.34	0.28									
Two-traits pedigree-based	Covariances matrix \mathbf{g}	1423.60 ± 0.87	45.58 ± 0.10	605.45 ± 0.42	21.79 ± 0.10	328.08 ± 0.22	26.25 ± 5.84E-02	25.93 ± 0.13	-0.34 ± 2.16E-02	8.66 ± 3.65E-02	-9.21E-02 ± 1.65E-02	0.38 ± 5.03E-03	0.10 ± 1.96E-03	0.37 ± 5.00E-03	-5.90E-03 ± 4.68E-03	0.45 ± 6.58E-03	-2.28E-02 ± 7.47E-03	0.10 ± 1.87E-03	-2.70E-02 ± 3.27E-03
	Covariances matrix \mathbf{p}	45.58 ± 0.10	23.61 ± 0.11	21.79 ± 0.10	24.36 ± 0.12	26.25 ± 5.84E-02	24.56 ± 0.12	-0.34 ± 2.16E-02	0.24 ± 1.19E-02	-9.21E-02 ± 1.65E-02	0.24 ± 9.51E-03	0.10 ± 1.96E-03	0.25 ± 2.91E-03	-5.90E-03 ± 4.68E-03	0.21 ± 7.78E-03	-2.28E-02 ± 7.47E-03	2.08E-01 ± 5.38E-03	-2.70E-02 ± 3.27E-03	0.20 ± 9.70E-03
	Covariances matrix \mathbf{E}	979.00 ± 0.54	0	310.71 ± 0.26	0	168.14 ± 0.14	0	4.91 ± 7.65E-02	0	1.28 ± 2.04E-02	0	0.32 ± 4.25E-03	0						
	h^2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Two-traits genomic-based	Covariances matrix \mathbf{g}	1432.00 ± 0.15	0.37 ± 9.63E-02	716.34 ± 0.13	-0.17 ± 3.99E-02	342.98 ± 6.08E-02	-0.30 ± 2.95E-02	21.27 ± 3.58E-02	5.62E-03 ± 6.19E-03	4.13 ± 7.10E-03	-6.10E-04 ± 3.05E-03	1.45 ± 3.08E-03	1.90E-03 ± 3.13E-03	0.56 ± 3.42E-03	-2.88E-03 ± 9.35E-03	0.85 ± 4.71E-03	-1.06E-02 ± 7.11E-03	0.24 ± 1.38E-03	-8.25E-04 ± 6.00E-03
	Covariances matrix \mathbf{p}	0.37 ± 9.63E-02	0.66 ± 7.21E-02	-0.17 ± 3.99E-02	0.76 ± 7.90E-02	-0.30 ± 2.95E-02	0.72 ± 7.37E-02	5.62E-03 ± 6.19E-03	7.43E-03 ± 7.75E-03	-6.10E-04 ± 3.05E-03	5.39E-03 ± 5.52E-03	1.90E-03 ± 3.13E-03	1.53E-03 ± 2.46E-03	-2.88E-03 ± 9.35E-03	3.67E-02 ± 4.58E-03	-1.06E-02 ± 7.11E-03	3.55E-02 ± 2.97E-03	-8.25E-04 ± 6.00E-03	3.88E-02 ± 5.71E-03
	Covariances matrix \mathbf{E}	979.00 ± 0.54	0	310.71 ± 0.26	0	168.14 ± 0.14	0	4.91 ± 7.65E-02	0	1.28 ± 2.04E-02	0	0.32 ± 4.25E-03	0						
	h^2	0.37	0.37	0.37	0.37	0.39	0.50	0.62	0.62	0.50	0.62	0.18	0.18	0.40	0.40	0.35	0.35	0.28	0.28

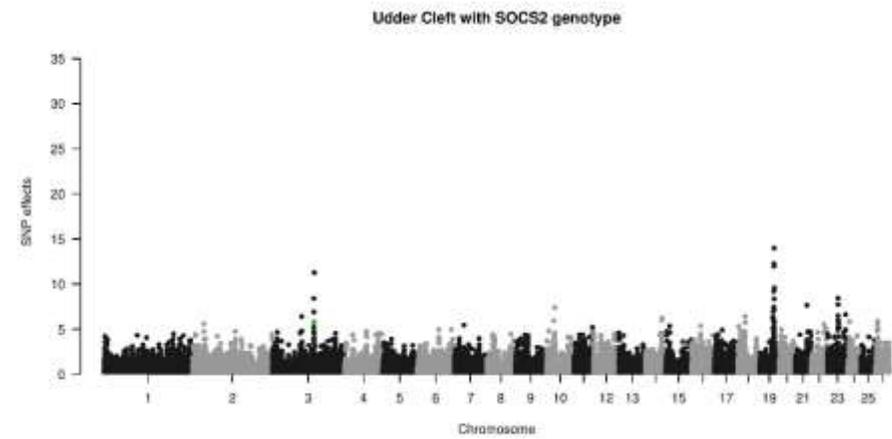
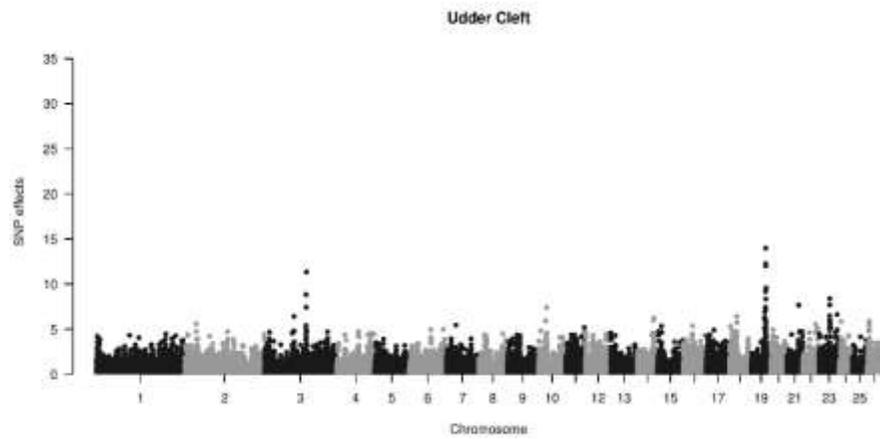
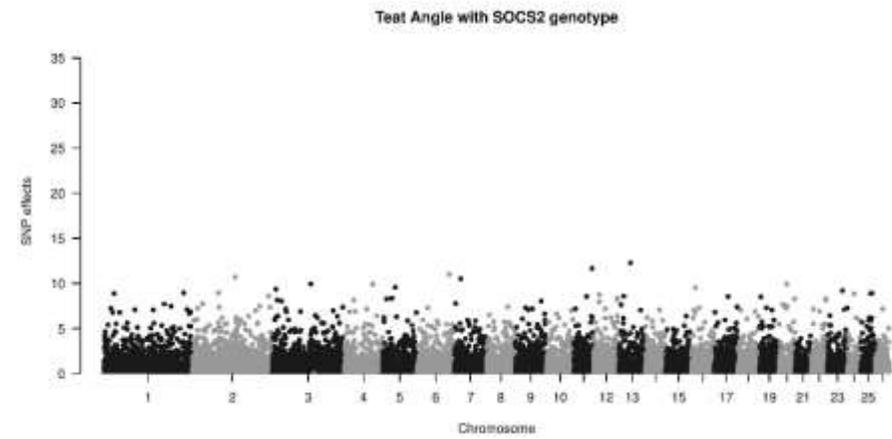
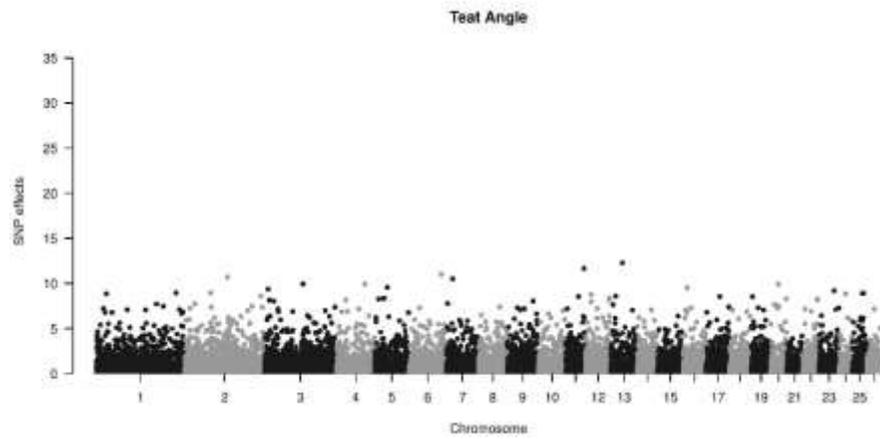
Abbreviations: GC: Gene Content; Gen. SD: Genetic standard deviation between homozygous individuals for the *SOCS2* genotype; MY: Milk Yield; FY: Fat Yield; PY: Protein Yield; FC: Fat Content; PC: Protein Content; SCS: Somatic Cell Score; TA: Teat Angle; UC: Udder Cleft; UD: Udder Depth

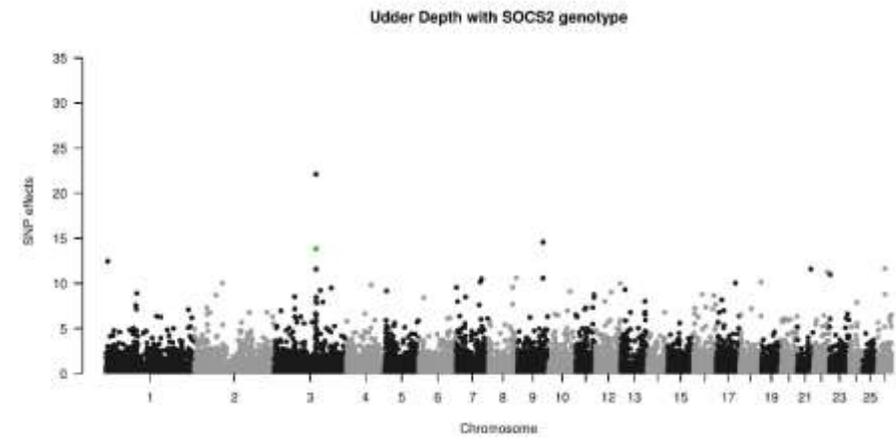
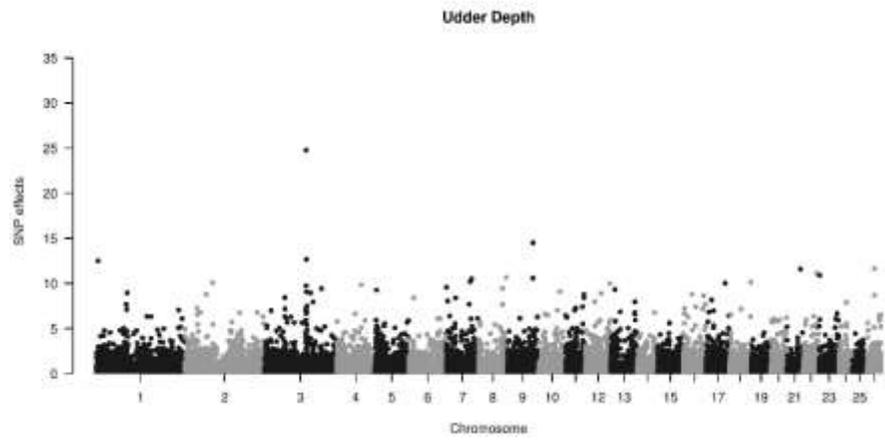
Annexe 5 Figure additionnelle (Figure S4) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique ([Chapitre II.](#)) : « Manhattan plots of estimated SNP effects using the best WssGBLUP approach for each phenotype (second iteration). On the left are presented analysis without the SOCS2 genotype among the markers and on the right, with the SOCS2 genotype (green point). »



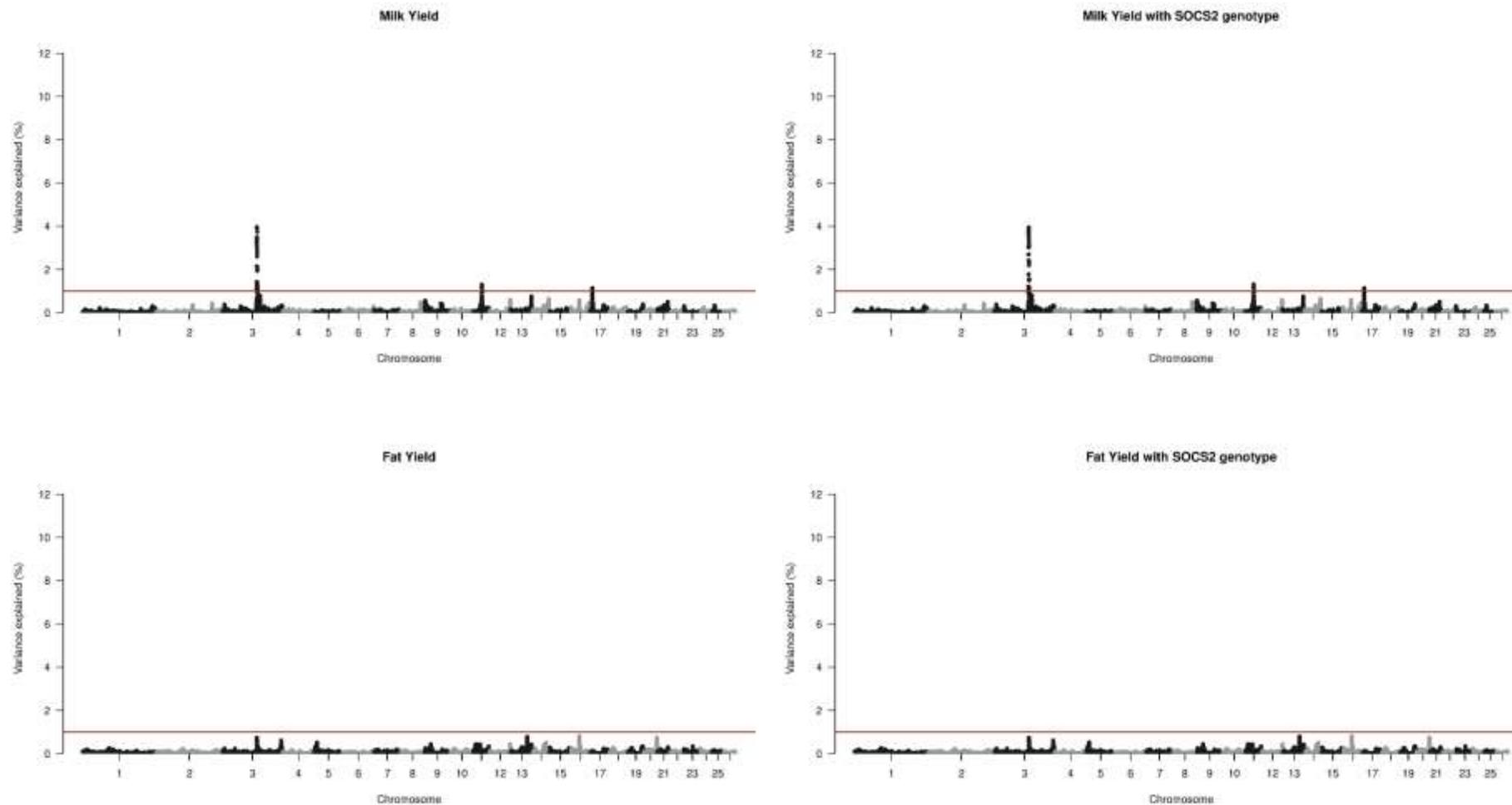


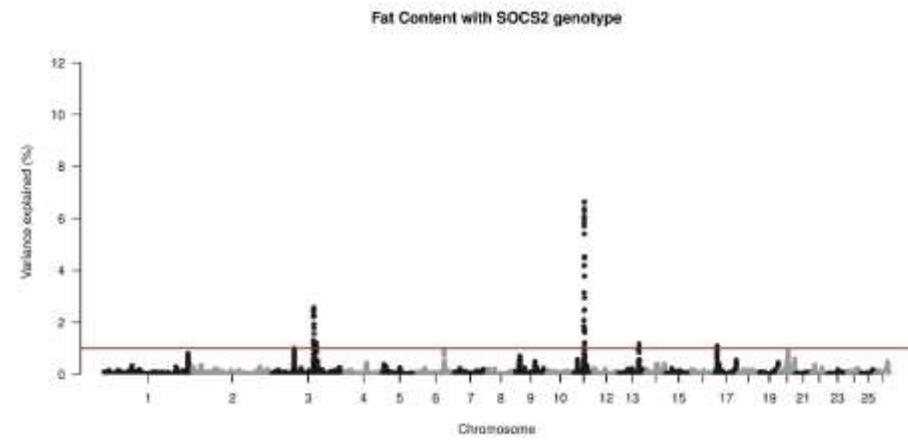
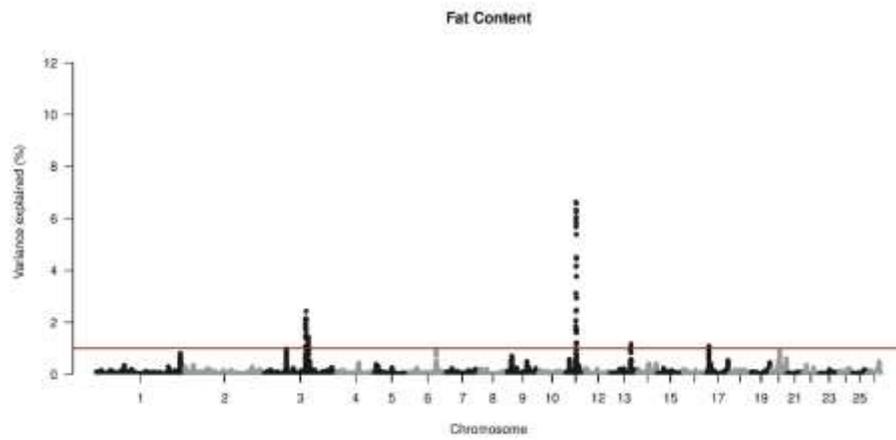
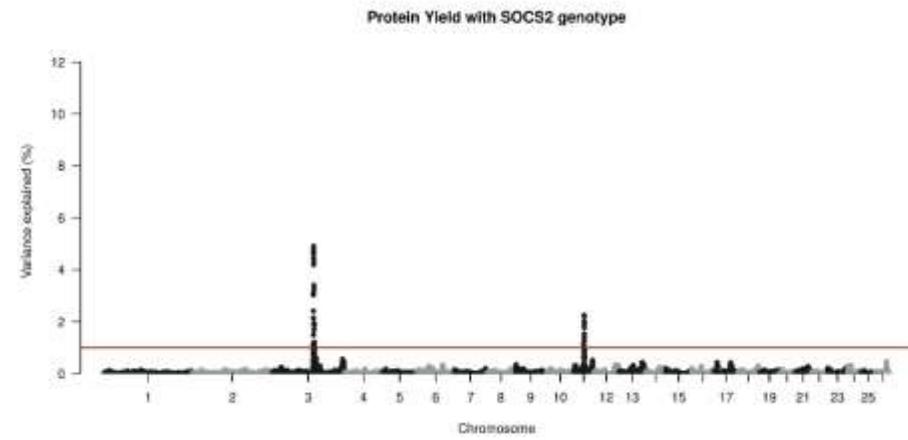
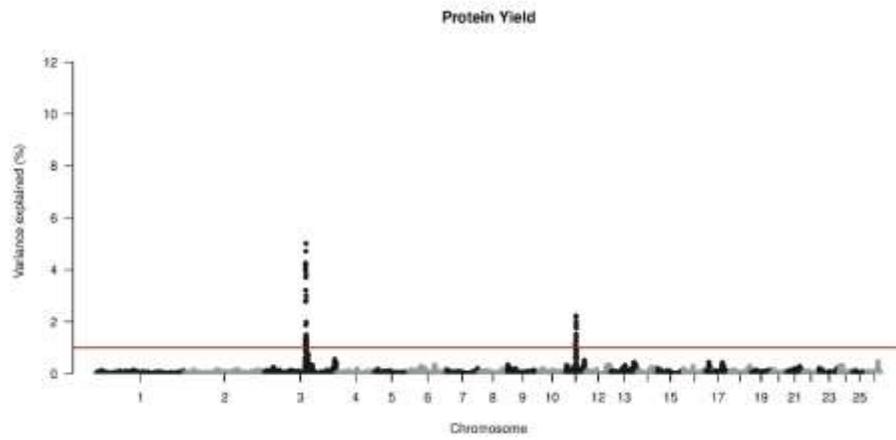


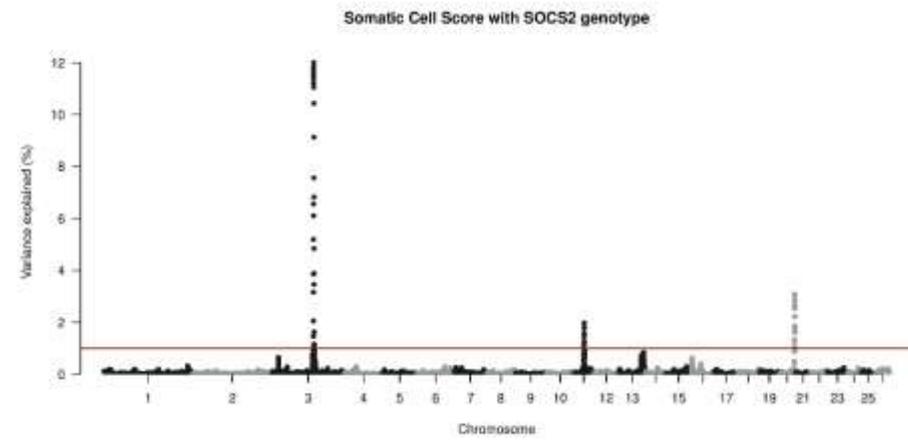
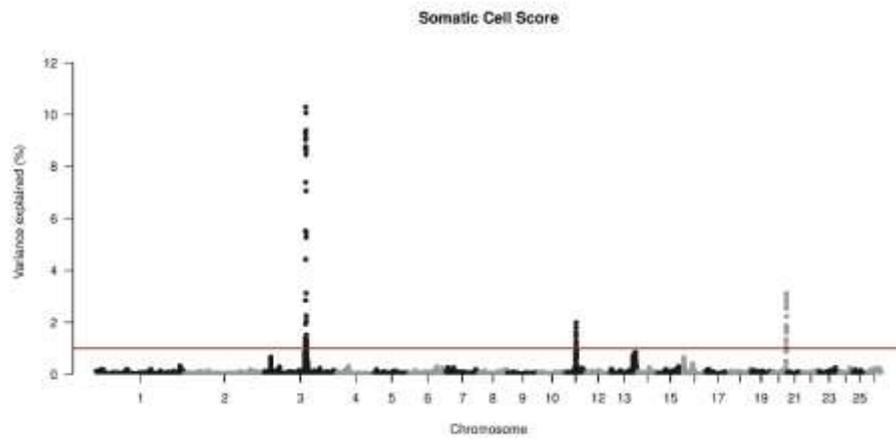
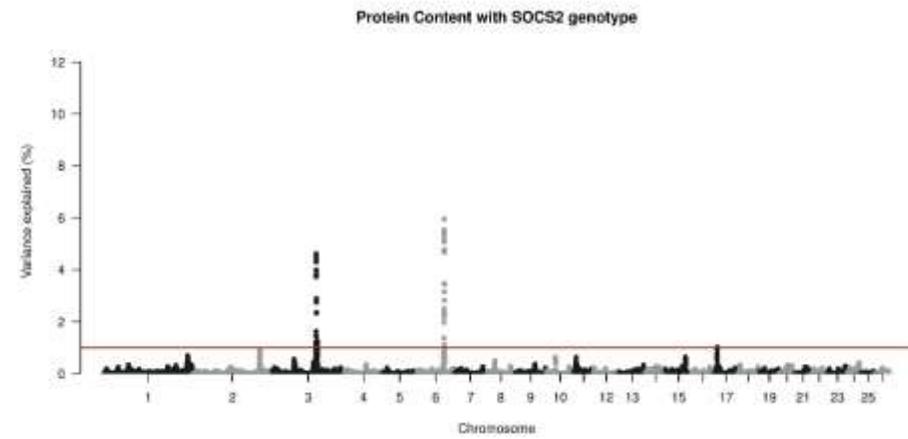
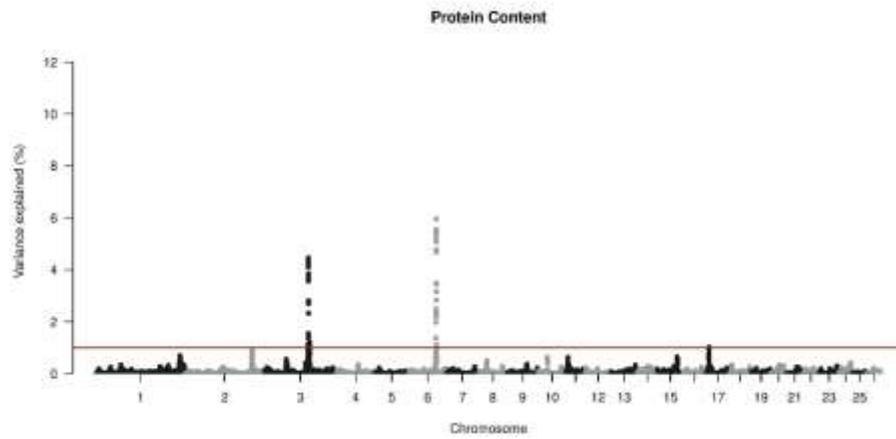




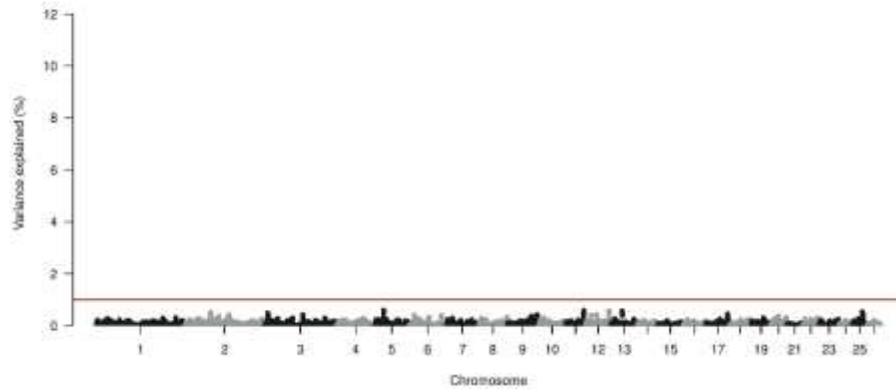
Annexe 6 Figure additionnelle (Figure S4) de l'article scientifique sur la prise en compte de la mutation SOCS2 dans le contexte de la sélection génomique ([Chapitre II](#)) : « Manhattan plots of estimated variance explained by 20 adjacent SNPs using the best WssGBLUP approach for each phenotype (second iteration). The horizontal red line represents the threshold of 1% adopted in this study. On the left are presented the analyses without the SOCS2 genotype among the markers and on the right, with the SOCS2 genotype. »



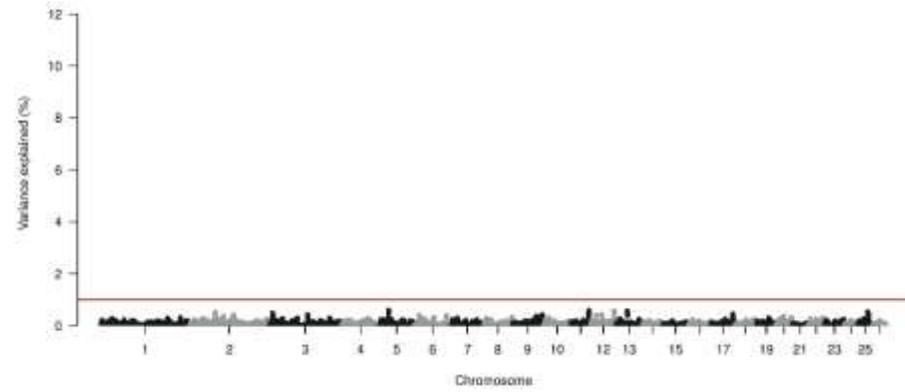




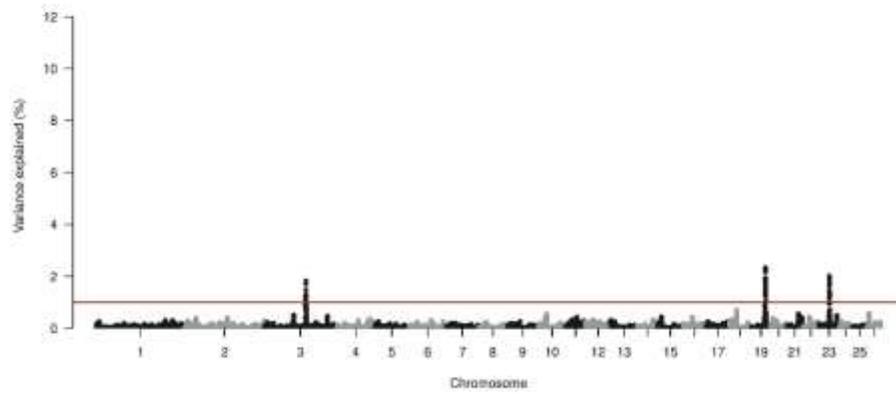
Teat Angle



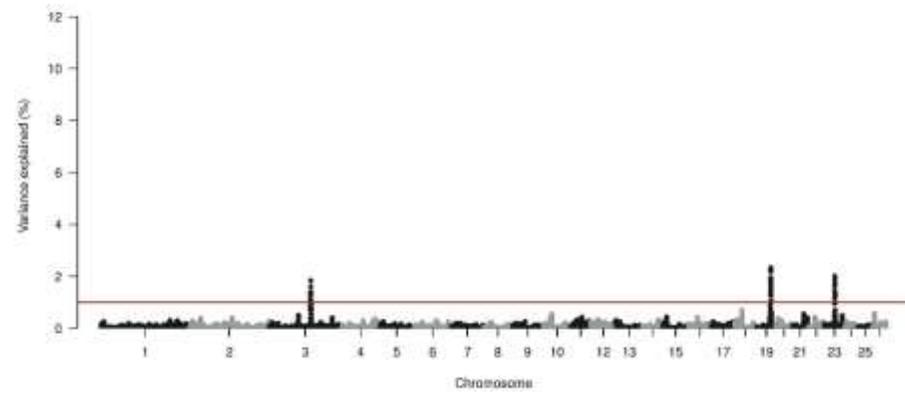
Teat Angle with SOCS2 genotype

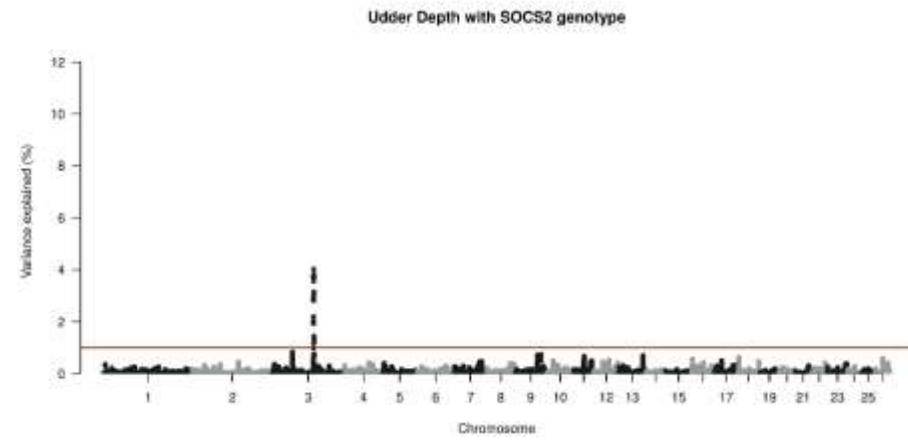
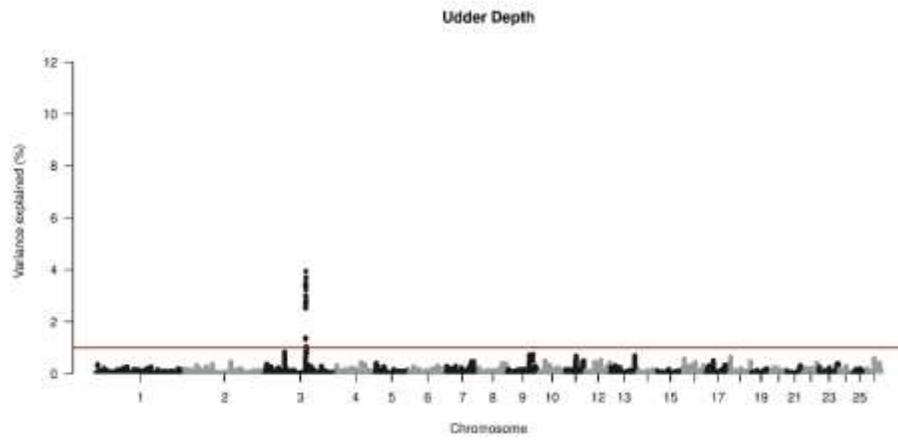


Udder Cleft



Udder Cleft with SOCS2 genotype

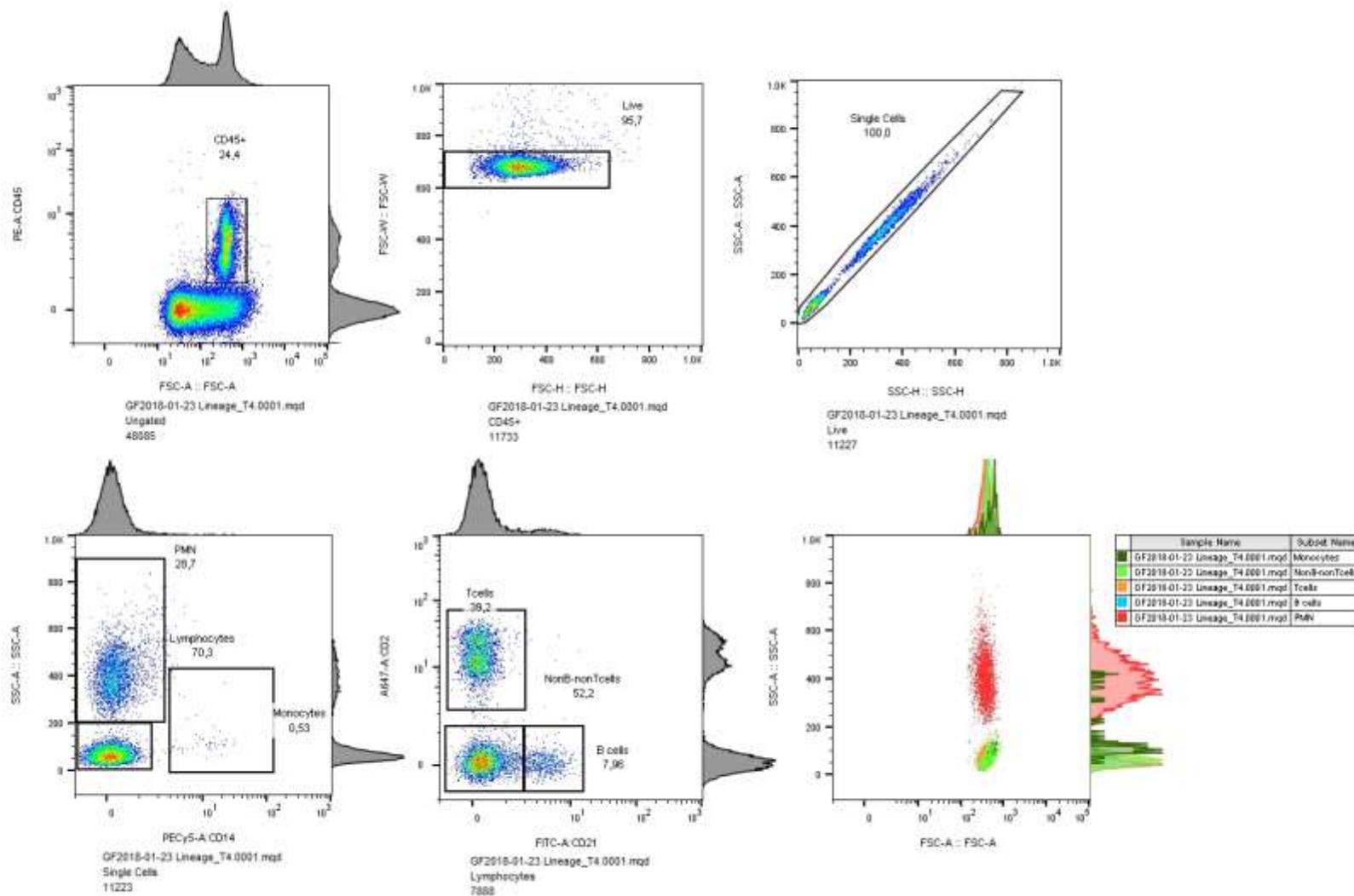




Annexe 7 Récapitulatif de la disponibilité des données générées au cours de l'infection expérimentale ([Chapitre IV.](#))

TEMPS	DONNEES CLINIQUES		DONNEES LAIT			DONNEES SANG						DONNEES AUTOPSIES	
	Thermobolus	Examen clinique	Bactériologie	Cellules somatiques	Production laitière	Typage leucocytes	Expression ARN	Dosage protéine inflammatoire	Typage cellules	Typage détaillé	Typage lymphocytes T	Dosage cytokines	Pesées et mesures
0	x	x	x	x	x	x	x	x	x	x	x	x	
4	x	x	x	x					x				
8	x	x	x	x	x		x	x	x			x	
12	x	x	x	x		x	x	x	x	x		x	
16	x	x	x	x	x		x	x	x			x	
20	x	x	x	x					x				
24	x	x	x	x		x	x	x	x	x	x	x	
28	x	x	x	x					x				
32	x	x	x	x	x				x				
36	x	x	x	x					x				
40	x	x	x	x	x				x				
44	x	x	x	x					(1 NA: 397)				
56	x	x	x	x	x	x	x	x	x	x	x	x	
64	x	x	x	x	x				x				
80	x	x	x	x	x				x				
88	x	x	x	x	x			x					
104	x	x	x	x	x		x	x	x			x	
128	x	x		x	x								
152	x (NA: 472)	x (NA: 472)	x	x (NA: 268 472)	x (NA: 472 inoc et non inoc)	x (NA: 268 472)			x (NA: 472)	x (NA: 472)	x (NA: 472)		
312													x

Annexe 8 Caractérisation des lymphocytes du sang avec le logiciel FlowJo® (Sarah Walachowski, [Chapitre IV.](#))



Annexe 9 Pipeline du traitement bioinformatique des données de séquence d'ARN de l'infection expérimentale ([Chapitre IV.](#))



Annexe 10 Résultats des modèles mixtes linéaires effectués sur les phénotypes de l'infection expérimentale ([Chapitre IV.](#))

Variable	Temps																			
	P-valeur (effet SOCS2)	C/C lsmeans ± erreur standard																		
		T/T lsmeans ± erreur standard																		
		P-valeur (effet SOCS2 intra-temps)																		
	0	4	8	12	16	20	24	28	32	36	40	44	56	64	80	88	104	128	152	
Température rumen imputée	39,3	39,4	39,7	39,7	41,3	40,8	40,5	40,2	39,8	39,7	39,5	39,3	39,2	39,3	40,8	40,9	40,5	40,1	40	
	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	
0,356	39,3	39,5	39,5	39,9	41,2	40,6	40,5	40,3	40,2	40,4	40,4	40,2	40	40,5	40,4	41,1	40,4	40,2	40,4	
	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	± 0,3	
Fréquence cardiaque	0,961	0,836	0,698	0,746	0,710	0,587	0,906	0,938	0,353	0,135	0,050	0,067	0,067	0,018	0,437	0,737	0,719	0,838	0,417	
	90,6	87,4	89,1	90	109,7	118,3	115,4	104,6	92,1	100,3	103,4	92	85,1	99,4	120,9	118,3	113,1	115,7	105	
0,926	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,6	
	84,9	91,7	89,7	90,9	103,4	106	107,4	110,6	103,7	98,6	98,9	96	89,1	104,6	110,6	120,9	114	108,9	111,4	
Fréquence respiratoire	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	± 7,2	
	0,583	0,680	0,956	0,934	0,547	0,249	0,445	0,565	0,276	0,869	0,660	0,700	0,700	0,621	0,330	0,804	0,934	0,511	0,550	
0,553	37,4	32,3	34	31,7	42,9	44	38,4	35,1	31,1	35,1	34,9	34,3	42,9	35,7	34,3	38,6	35,1	32,1	26,2	
	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,5	
Score clinique systémique	34,9	30,9	39,4	33,4	38,3	43,4	39,4	35,1	29,1	36	36,6	33,1	33,7	35,7	30	32,6	32	27,1	27,4	
	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	± 3,3	
0,891	0,596	0,767	0,272	0,723	0,352	0,906	0,836	1,000	0,679	0,859	0,723	0,813	0,077	1,000	0,382	0,228	0,518	0,310	0,800	
	1	0,7	0,9	0,8	4,9	4,4	3,3	2	1,6	0,9	2	0,7	1,6	0,9	2,8	2,9	3,2	2,2	2,2	
Score clinique mammaire inoc.	± 0,5	± 0,4	± 0,5	± 0,4	± 1,4	± 1,3	± 1,0	± 0,7	± 0,6	± 0,5	± 0,7	± 0,4	± 0,6	± 0,5	± 0,9	± 0,9	± 1,0	± 0,8	± 0,8	
	0,2	0,7	1,3	0,5	3,2	3,2	1,9	2,3	1,6	1,9	1,4	0,9	1,4	2,4	2,3	2,8	3,2	3,1	1,8	
0,180	± 0,3	± 0,4	± 0,5	± 0,4	± 1,0	± 1,0	± 0,7	± 0,8	± 0,6	± 0,7	± 0,6	± 0,5	± 0,6	± 0,8	± 0,8	± 0,9	± 1,0	± 1,0	± 0,7	
	0,9	1,2	1,5	1,9	1,6	2,5	2,9	2,5	2,7	2	3,4	3,9	3,4	2,5	3,8	4,5	5,8	4,9	6,1	
0,9	± 0,2	± 0,3	± 0,3	± 0,4	± 0,3	± 0,4	± 0,5	± 0,4	± 0,5	± 0,4	± 0,6	± 0,6	± 0,6	± 0,4	± 0,6	± 0,7	± 0,9	± 0,8	± 1,0	
	1	1,8	1,9	1,9	2,2	3,1	2,5	3,3	3,6	2,9	3,4	4	5,2	5,1	4,3	4,8	5	5,4	6,3	
0,180	± 0,3	± 0,4	± 0,4	± 0,4	± 0,4	± 0,5	± 0,5	± 0,5	± 0,6	± 0,5	± 0,6	± 0,6	± 0,8	± 0,8	± 0,7	± 0,7	± 0,8	± 0,8	± 0,9	

0,072	0,716	0,193	0,403	0,935	0,290	0,373	0,573	0,286	0,249	0,164	0,989	0,944	0,087	0,013	0,613	0,842	0,512	0,628	0,868
<i>Score clinique mammaire non inoc.</i>	1 ± 0,3	1,6 ± 0,4	1,6 ± 0,4	1,6 ± 0,4	1,4 ± 0,4	1,8 ± 0,5	1,7 ± 0,5	1,1 ± 0,4	1,3 ± 0,4	1,9 ± 0,5	2,1 ± 0,5	2,5 ± 0,6	2,3 ± 0,6	1,9 ± 0,5	1,9 ± 0,5	1,8 ± 0,5	2,3 ± 0,6	1,8 ± 0,5	2,3 ± 0,6
0,613	1,1 ± 0,4	1,6 ± 0,4	1,9 ± 0,5	1,7 ± 0,5	1,4 ± 0,4	2,2 ± 0,5	2,4 ± 0,6	1,7 ± 0,5	1,6 ± 0,5	1,9 ± 0,5	1,6 ± 0,4	2,5 ± 0,6	2,6 ± 0,6	2,8 ± 0,6	2,4 ± 0,6	2,2 ± 0,5	2,9 ± 0,7	1,9 ± 0,5	2,3 ± 0,6
	0,885	0,916	0,647	0,897	0,927	0,616	0,352	0,320	0,559	0,993	0,417	0,981	0,699	0,295	0,531	0,550	0,493	0,870	0,989
<i>Bactériologie mamelle inoc.</i>	7,9E+00 ± 1,4E+01	1,7E+05 ± 2,6E+05	3,6E+07 ± 5,5E+07	1,7E+07 ± 2,6E+07	4,5E+06 ± 6,7E+06	2,3E+06 ± 3,4E+06	3,0E+06 ± 4,5E+06	6,8E+05 ± 1,0E+06	3,8E+05 ± 5,8E+05	1,6E+05 ± 2,5E+05	1,3E+06 ± 2,0E+06	2,3E+06 ± 3,4E+06	1,1E+06 ± 1,7E+06	1,3E+06 ± 1,9E+06	3,0E+07 ± 4,6E+07	6,7E+04 ± 1,0E+05	2,2E+06 ± 3,3E+06		2,4E+06 ± 3,7E+06
0,407	1,3E+01 ± 2,2E+01	6,6E+03 ± 1,0E+04	1,3E+07 ± 1,9E+07	6,5E+06 ± 9,8E+06	5,5E+06 ± 8,4E+06	2,3E+06 ± 3,5E+06	3,0E+06 ± 4,6E+06	1,8E+06 ± 2,8E+06	9,4E+04 ± 1,4E+05	9,1E+05 ± 1,4E+06	1,5E+06 ± 2,3E+06	5,2E+05 ± 7,8E+05	9,1E+04 ± 1,4E+05	1,1E+06 ± 1,6E+06	1,4E+06 ± 2,1E+06	7,3E+06 ± 1,1E+07	4,4E+05 ± 6,7E+05		7,7E+04 ± 1,2E+05
	0,839	0,538	0,692	0,712	0,922	0,990	0,996	0,704	0,637	0,603	0,963	0,629	0,557	0,939	0,541	0,525	0,616		0,535
<i>Score cellules somatiques mamelle inoc.</i>	1,5 ± 0,5	2,5 ± 0,5	2,5 ± 0,5	7,4 ± 0,5	12,0 ± 0,5	13,0 ± 0,5	13,4 ± 0,5	12,8 ± 0,5	12,7 ± 0,5	12,0 ± 0,5	11,4 ± 0,5	10,2 ± 0,5	9,0 ± 0,5	7,5 ± 0,5	9,8 ± 0,5	11,6 ± 0,5	11,4 ± 0,5	10,0 ± 0,5	9,6 ± 0,5
0,256	2,7 ± 0,5	3,5 ± 0,5	3,3 ± 0,5	7,7 ± 0,5	11,4 ± 0,5	12,7 ± 0,5	12,2 ± 0,5	11,9 ± 0,5	11,8 ± 0,5	11,4 ± 0,5	10,8 ± 0,5	9,8 ± 0,5	8,5 ± 0,5	8,1 ± 0,5	9,0 ± 0,5	10,3 ± 0,5	10,9 ± 0,5	9,6 ± 0,5	9,3 ± 0,5
	0,102	0,161	0,261	0,637	0,378	0,674	0,088	0,196	0,192	0,332	0,368	0,519	0,411	0,342	0,245	0,069	0,479	0,518	0,720
<i>Production laitière</i>	1164,3 ± 114,5		1097,1 ± 114,5		542,9 ± 114,5				724,3 ± 114,5		415,7 ± 114,5		870,0 ± 114,5	570,0 ± 114,5	857,1 ± 114,5	408,6 ± 114,5	597,1 ± 114,5	622,9 ± 114,5	692,8 ± 120,0
0,931	1065,7 ± 114,5		1191,4 ± 114,5		588,6 ± 114,5				880,0 ± 114,5		497,1 ± 114,5		907,1 ± 114,5	495,7 ± 114,5	654,3 ± 114,5	471,4 ± 114,5	554,3 ± 114,5	605,7 ± 114,5	532,9 ± 114,5
	0,554		0,571		0,783				0,355		0,624		0,823	0,655	0,234	0,705	0,796	0,368	0,354
<i>Production laitière mamelle inoc.</i>	617,1 ± 52,1		591,4 ± 52,1		274,3 ± 52,1				255,7 ± 52,1		132,9 ± 52,1		355,7 ± 52,1	241,4 ± 52,1	397,1 ± 52,1	125,7 ± 52,1	188,6 ± 52,1	237,1 ± 52,1	308,5 ± 55,4
0,682	592,9 ± 52,1		640,0 ± 52,1		294,3 ± 52,1				355,7 ± 52,1		202,9 ± 52,1		407,1 ± 52,1	235,7 ± 52,1	322,9 ± 52,1	165,7 ± 52,1	232,9 ± 52,1	245,7 ± 52,1	247,1 ± 52,1
	0,748		0,522		0,791				0,200		0,361		0,499	0,940	0,333	0,597	0,559	0,368	0,436
<i>Production laitière mamelle non inoc.</i>	547,1 ± 83,5		505,7 ± 83,5		268,6 ± 83,5				468,6 ± 83,5		282,9 ± 83,5		514,3 ± 83,5	328,6 ± 83,5	460,0 ± 83,5	282,9 ± 83,5	408,6 ± 83,5	385,7 ± 83,5	384,4 ± 86,4
0,770	472,9 ± 83,5		551,4 ± 83,5		294,3 ± 83,5				524,3 ± 83,5		294,3 ± 83,5		500,0 ± 83,5	260,0 ± 83,5	331,4 ± 83,5	305,7 ± 83,5	321,4 ± 83,5	360,0 ± 83,5	285,7 ± 83,5
	0,541		0,705		0,831				0,645		0,925		0,906	0,572	0,297	0,850	0,475	0,368	0,427

<i>Neutrophiles vivants</i>	8,5E+02 ± 5,0E+02		1,3E+06 ± 7,7E+05		1,9E+07 ± 1,1E+07													4,6E+06 ± 2,7E+06		1,7E+06 ± 1,1E+06	
<i>mamelle inoc. 0,861</i>	3,6E+03 ± 2,1E+03		1,4E+06 ± 8,0E+05		1,1E+07 ± 6,4E+06													2,1E+06 ± 1,2E+06		2,4E+06 ± 1,5E+06	
	0,231		0,967		0,540													0,416		0,733	
<i>Neutrophiles morts mamelle inoc. 0,776</i>	2,4E+04 ± 1,1E+04		1,8E+05 ± 8,4E+04		2,0E+07 ± 9,0E+06													1,1E+06 ± 4,8E+05		2,1E+06 ± 1,0E+06	
	5,2E+04 ± 2,4E+04		1,8E+05 ± 8,4E+04		1,3E+07 ± 5,9E+06													9,2E+05 ± 4,2E+05		2,6E+06 ± 1,3E+06	
	0,311		0,992		0,542													0,840		0,749	
<i>Monocytes et macrophages mamelle inoc. 0,577</i>	1,8E+03 ± 8,1E+02		1,9E+05 ± 8,3E+04		5,9E+07 ± 2,6E+07													5,1E+05 ± 2,2E+05		4,9E+05 ± 2,3E+05	
	3,1E+03 ± 1,4E+03		2,6E+05 ± 1,1E+05		1,9E+07 ± 8,4E+06													2,6E+05 ± 1,1E+05		5,7E+05 ± 2,7E+05	
	0,429		0,624		0,172													0,343		0,808	
<i>Lymphocytes mamelle inoc. 0,298</i>	6,8E+03 ± 2,7E+03		6,0E+04 ± 2,4E+04		6,5E+06 ± 2,6E+06													1,9E+05 ± 7,7E+04		2,9E+05 ± 1,3E+05	
	1,2E+04 ± 5,0E+03		7,5E+04 ± 3,0E+04		1,1E+06 ± 4,3E+05													1,4E+05 ± 5,8E+04		2,5E+05 ± 1,1E+05	
	0,352		0,696		0,063													0,625		0,781	
<i>Haptoglobine 0,125</i>	0,66 ± 0,2	5,95 ± 1,0	5,99 ± 1,0	6,33 ± 1,1		18,24 ± 2,9												6,63 ± 1,1		6,18 ± 1,1	8,59 ± 1,4
	0,48 ± 0,2	6,14 ± 1,1	5,74 ± 1,0	6,67 ± 1,1		20,46 ± 3,2												8,62 ± 1,4		16,18 ± 2,6	10,20 ± 1,7
	0,587	0,905	0,868	0,833		0,616												0,300		0,004	0,480
<i>Leucocytes 0,952</i>	6,1E+06 ± 7,6E+05	5,9E+06 ± 7,4E+05	5,9E+06 ± 7,4E+05	5,5E+06 ± 6,9E+05	4,1E+06 ± 5,1E+05	4,7E+06 ± 5,9E+05	5,8E+06 ± 7,3E+05	7,0E+06 ± 8,8E+05	7,4E+06 ± 9,2E+05	7,6E+06 ± 9,5E+05	7,2E+06 ± 9,0E+05	7,2E+06 ± 9,1E+05	5,9E+06 ± 7,3E+05	5,6E+06 ± 7,0E+05	4,0E+06 ± 5,0E+05					4,3E+06 ± 5,4E+05	3,8E+06 ± 4,9E+05
	5,7E+06 ± 7,2E+05	6,1E+06 ± 7,6E+05	6,0E+06 ± 7,5E+05	6,1E+06 ± 7,6E+05	4,9E+06 ± 6,1E+05	5,6E+06 ± 7,0E+05	5,3E+06 ± 6,6E+05	6,7E+06 ± 8,3E+05	7,0E+06 ± 8,8E+05	6,8E+06 ± 8,5E+05	7,1E+06 ± 8,9E+05	6,9E+06 ± 8,7E+05	6,4E+06 ± 8,0E+05	5,6E+06 ± 7,1E+05	4,3E+06 ± 5,4E+05					4,2E+06 ± 5,3E+05	3,8E+06 ± 4,7E+05
	0,745	0,850	0,968	0,617	0,323	0,357	0,616	0,776	0,788	0,535	0,974	0,803	0,614	0,970	0,685					0,861	0,972

<i>Monocytes</i> <i>0,443</i>	2,9E+04	5,6E+04	4,4E+04	2,7E+04	2,7E+04	5,7E+04	3,7E+04	5,0E+04	4,1E+04	4,3E+04	4,8E+04	3,2E+04	4,2E+04	9,9E+04	1,3E+04	1,7E+04	2,0E+04
	± 1,1E+04	± 2,1E+04	± 1,7E+04	± 1,0E+04	± 1,0E+04	± 2,1E+04	± 1,4E+04	± 1,9E+04	± 1,5E+04	± 1,6E+04	± 1,8E+04	± 1,2E+04	± 1,6E+04	± 3,8E+04	± 4,9E+03	± 6,5E+03	± 7,6E+03
	2,1E+04	3,7E+04	2,8E+04	2,0E+04	2,2E+04	3,1E+04	2,0E+04	2,3E+04	2,3E+04	2,6E+04	3,5E+04	3,3E+04	3,7E+04	6,8E+04	2,0E+04	9,6E+03	1,0E+04
± 7,8E+03	± 1,4E+04	± 1,1E+04	± 7,7E+03	± 8,3E+03	± 1,2E+04	± 7,6E+03	± 8,7E+03	± 8,7E+03	± 9,9E+03	± 1,3E+04	± 1,3E+04	± 1,4E+04	± 2,6E+04	± 7,5E+03	± 3,6E+03	± 3,8E+03	
	0,525	0,464	0,453	0,626	0,725	0,308	0,309	0,225	0,338	0,383	0,566	0,956	0,799	0,499	0,454	0,335	0,283
<i>Neutrophiles</i> <i>0,933</i>	2,1E+06	1,8E+06	1,9E+06	1,5E+06	1,7E+06	2,4E+06	3,0E+06	3,5E+06	3,6E+06	3,3E+06	2,8E+06	2,6E+06	1,8E+06	1,7E+06	1,5E+06	1,3E+06	1,3E+06
	± 3,8E+05	± 3,2E+05	± 3,4E+05	± 2,7E+05	± 2,9E+05	± 4,3E+05	± 5,4E+05	± 6,3E+05	± 6,3E+05	± 5,9E+05	± 5,1E+05	± 4,6E+05	± 3,3E+05	± 3,1E+05	± 2,7E+05	± 2,3E+05	± 2,5E+05
	1,8E+06	1,9E+06	1,9E+06	1,9E+06	2,0E+06	2,6E+06	2,4E+06	3,0E+06	3,2E+06	2,9E+06	2,9E+06	2,7E+06	2,4E+06	2,1E+06	1,6E+06	1,5E+06	1,2E+06
± 3,2E+05	± 3,4E+05	± 3,4E+05	± 3,5E+05	± 3,6E+05	± 4,7E+05	± 4,2E+05	± 5,4E+05	± 5,6E+05	± 5,2E+05	± 5,2E+05	± 5,0E+05	± 4,2E+05	± 3,8E+05	± 2,8E+05	± 2,6E+05	± 2,2E+05	
	0,487	0,765	0,935	0,386	0,413	0,738	0,343	0,562	0,645	0,637	0,888	0,903	0,323	0,454	0,879	0,677	0,750
<i>Lymphocytes</i> <i>0,924</i>	3,8E+06	3,9E+06	3,8E+06	3,9E+06	2,3E+06	2,2E+06	2,6E+06	3,3E+06	3,6E+06	4,2E+06	4,2E+06	4,5E+06	3,9E+06	3,7E+06	2,3E+06	2,9E+06	2,3E+06
	± 5,6E+05	± 5,7E+05	± 5,5E+05	± 5,7E+05	± 3,3E+05	± 3,2E+05	± 3,9E+05	± 4,8E+05	± 5,3E+05	± 6,1E+05	± 6,1E+05	± 6,6E+05	± 5,8E+05	± 5,4E+05	± 3,4E+05	± 4,2E+05	± 3,5E+05
	3,8E+06	3,9E+06	3,8E+06	3,8E+06	2,7E+06	2,8E+06	2,8E+06	3,4E+06	3,7E+06	3,6E+06	3,7E+06	3,8E+06	3,6E+06	3,2E+06	2,5E+06	2,6E+06	2,3E+06
± 5,5E+05	± 5,7E+05	± 5,6E+05	± 5,6E+05	± 3,9E+05	± 4,1E+05	± 4,1E+05	± 5,0E+05	± 5,4E+05	± 5,3E+05	± 5,4E+05	± 5,7E+05	± 5,3E+05	± 4,6E+05	± 3,7E+05	± 3,8E+05	± 3,3E+05	
	0,932	0,957	0,990	0,932	0,458	0,269	0,799	0,874	0,977	0,532	0,577	0,431	0,700	0,489	0,712	0,622	0,954
<i>Lymphocytes B</i> <i>0,418</i>	2,3E+05	2,7E+05	2,5E+05	2,8E+05	1,0E+05	1,2E+05	1,5E+05	2,7E+05	3,6E+05	4,4E+05	4,2E+05	4,3E+05	3,0E+05	2,8E+05	1,1E+05	1,9E+05	1,3E+05
	± 6,2E+04	± 7,2E+04	± 6,8E+04	± 7,6E+04	± 2,7E+04	± 3,3E+04	± 4,1E+04	± 7,3E+04	± 9,6E+04	± 1,2E+05	± 1,1E+05	± 1,2E+05	± 8,1E+04	± 7,6E+04	± 2,9E+04	± 5,0E+04	± 3,7E+04
	1,8E+05	2,2E+05	1,9E+05	1,6E+05	9,2E+04	1,3E+05	1,4E+05	1,9E+05	2,3E+05	2,6E+05	2,9E+05	2,5E+05	2,2E+05	1,9E+05	1,3E+05	1,2E+05	1,1E+05
± 4,8E+04	± 5,8E+04	± 5,1E+04	± 4,2E+04	± 2,5E+04	± 3,5E+04	± 3,9E+04	± 5,1E+04	± 6,2E+04	± 7,1E+04	± 7,8E+04	± 6,8E+04	± 5,8E+04	± 5,1E+04	± 3,6E+04	± 3,1E+04	± 3,0E+04	
	0,520	0,586	0,467	0,174	0,819	0,910	0,863	0,391	0,284	0,233	0,358	0,199	0,424	0,329	0,589	0,265	0,650
<i>Lymphocytes T</i> <i>0,335</i>	1,5E+06	1,5E+06	1,5E+06	1,6E+06	8,3E+05	8,0E+05	1,1E+06	1,4E+06	1,4E+06	1,7E+06	1,7E+06	1,8E+06	1,5E+06	1,5E+06	8,6E+05	1,2E+06	9,6E+05
	± 2,0E+05	± 2,1E+05	± 2,1E+05	± 2,2E+05	± 1,1E+05	± 1,1E+05	± 1,5E+05	± 1,9E+05	± 2,0E+05	± 2,3E+05	± 2,3E+05	± 2,5E+05	± 2,1E+05	± 2,0E+05	± 1,2E+05	± 1,6E+05	± 1,4E+05
	1,8E+06	1,9E+06	1,9E+06	1,9E+06	1,2E+06	1,3E+06	1,3E+06	1,7E+06	1,8E+06	1,7E+06	1,8E+06	1,7E+06	1,6E+06	1,4E+06	1,1E+06	1,2E+06	1,0E+06
± 2,5E+05	± 2,6E+05	± 2,5E+05	± 2,6E+05	± 1,7E+05	± 1,7E+05	± 1,8E+05	± 2,3E+05	± 2,4E+05	± 2,3E+05	± 2,4E+05	± 2,5E+05	± 2,2E+05	± 1,9E+05	± 1,5E+05	± 1,6E+05	± 1,4E+05	
	0,320	0,305	0,351	0,460	0,070	0,046	0,253	0,358	0,340	1,000	0,787	0,864	0,782	0,845	0,248	0,939	0,728
<i>Lymphocytes non B et non T</i> <i>0,587</i>	2,1E+06	2,0E+06	1,9E+06	1,8E+06	1,3E+06	1,2E+06	1,4E+06	1,6E+06	1,8E+06	2,0E+06	2,0E+06	2,2E+06	2,1E+06	1,9E+06	1,3E+06	1,5E+06	1,2E+06
	± 3,5E+05	± 3,4E+05	± 3,3E+05	± 3,1E+05	± 2,2E+05	± 2,1E+05	± 2,4E+05	± 2,8E+05	± 3,1E+05	± 3,4E+05	± 3,4E+05	± 3,7E+05	± 3,5E+05	± 3,2E+05	± 2,3E+05	± 2,5E+05	± 2,1E+05
	1,7E+06	1,7E+06	1,7E+06	1,7E+06	1,3E+06	1,3E+06	1,2E+06	1,5E+06	1,6E+06	1,6E+06	1,7E+06	1,7E+06	1,7E+06	1,5E+06	1,2E+06	1,2E+06	1,1E+06
± 2,9E+05	± 2,9E+05	± 2,9E+05	± 3,0E+05	± 2,2E+05	± 2,3E+05	± 2,1E+05	± 2,5E+05	± 2,8E+05	± 2,8E+05	± 3,0E+05	± 3,0E+05	± 3,0E+05	± 2,6E+05	± 2,1E+05	± 2,1E+05	± 1,9E+05	
	0,434	0,547	0,610	0,823	0,987	0,693	0,612	0,737	0,666	0,448	0,573	0,341	0,484	0,383	0,843	0,503	0,749

<i>Monocytes conventionnels</i>	3,0E+04 ± 1,2E+04	2,6E+04 ± 1,1E+04	5,2E+04 ± 2,1E+04	4,2E+04 ± 1,7E+04	2,9E+04 ± 1,2E+04
0,584	2,2E+04 ± 9,1E+03	2,4E+04 ± 9,9E+03	2,1E+04 ± 8,5E+03	5,3E+04 ± 2,2E+04	1,8E+04 ± 7,5E+03
	0,638	0,883	0,198	0,686	0,478
<i>Monocytes inflammatoires</i>	2,9E+04 ± 1,1E+04	2,8E+04 ± 1,0E+04	2,3E+04 ± 8,6E+03	3,9E+04 ± 1,4E+04	2,0E+04 ± 7,5E+03
0,841	2,8E+04 ± 1,0E+04	2,8E+04 ± 1,0E+04	2,5E+04 ± 9,1E+03	4,0E+04 ± 1,4E+04	1,3E+04 ± 4,6E+03
	0,935	0,989	0,909	0,970	0,414
<i>Monocytes intermédiaires</i>	1,1E+04 ± 4,0E+03	6,9E+03 ± 2,5E+03	8,3E+03 ± 3,0E+03	1,5E+04 ± 5,3E+03	7,8E+03 ± 3,0E+03
0,407	6,7E+03 ± 2,4E+03	4,5E+03 ± 1,6E+03	5,4E+03 ± 2,0E+03	1,2E+04 ± 4,3E+03	6,5E+03 ± 2,3E+03
	0,379	0,449	0,433	0,685	0,732
<i>Lymphocytes NK</i>	6,3E+04 ± 2,2E+04	4,8E+04 ± 1,7E+04	5,1E+04 ± 1,8E+04	5,4E+04 ± 1,9E+04	4,2E+04 ± 1,5E+04
0,329	4,6E+04 ± 1,6E+04	3,9E+04 ± 1,4E+04	3,9E+04 ± 1,4E+04	4,5E+04 ± 1,6E+04	1,2E+04 ± 4,3E+03
	0,531	0,701	0,582	0,696	0,080
<i>Lymphocytes non NK</i>	3,8E+06 ± 5,9E+05	3,8E+06 ± 5,8E+05	2,5E+06 ± 3,9E+05	3,8E+06 ± 6,0E+05	2,2E+06 ± 3,5E+05
0,772	3,7E+06 ± 5,7E+05	3,7E+06 ± 5,7E+05	2,5E+06 ± 3,9E+05	3,5E+06 ± 5,5E+05	2,0E+06 ± 3,2E+05
	0,880	0,939	0,955	0,708	0,801
<i>Ratio neutrophiles et monocytes inflammatoires</i>	39.1 ± 12.7	35.3 ± 11.5	42.2 ± 13.7	29.3 ± 9.6	34.5 ± 11.5
0,279	45.9 ± 14.8	62.2 ± 20.0	57.3 ± 18.4	47.6 ± 15.4	78.7 ± 25.2
	0,734	0,267	0,523	0,331	0,137

<i>PBMC</i>	4,0E+06	3,9E+06	2,7E+06	4,0E+06	2,3E+06
<i>0,769</i>	± 5,8E+05	± 5,7E+05	± 3,9E+05	± 5,8E+05	± 3,5E+05
	3,9E+06	3,8E+06	2,6E+06	3,8E+06	2,2E+06
	± 5,6E+05	± 5,6E+05	± 3,8E+05	± 5,5E+05	± 3,1E+05
	0,867	0,955	0,920	0,801	0,739
<i>Lymphocytes T</i>	9,8E+05		6,6E+05	1,0E+06	6,4E+05
<i>CD4+</i>	± 1,5E+05		± 9,9E+04	± 1,5E+05	± 1,0E+05
<i>0,784</i>	1,1E+06		8,0E+05	9,8E+05	6,2E+05
	± 1,6E+05		± 1,2E+05	± 1,5E+05	± 9,3E+04
	0,636		0,383	0,790	0,883
<i>Lymphocytes T</i>	4,0E+05		3,4E+05	4,1E+05	2,4E+05
<i>CD8+</i>	± 5,8E+04		± 4,9E+04	± 5,9E+04	± 3,8E+04
<i>0,073</i>	6,0E+05		4,6E+05	5,4E+05	3,0E+05
	± 8,6E+04		± 6,6E+04	± 7,7E+04	± 4,3E+04
	0,086		0,183	0,217	0,364
<i>Lymphocytes T</i>	8,3E+04		5,1E+04	8,7E+04	7,1E+04
<i>CD4- et CD8-</i>	± 2,0E+04		± 1,2E+04	± 2,1E+04	± 1,8E+04
<i>0,377</i>	1,2E+05		7,5E+04	9,8E+04	7,7E+04
	± 2,7E+04		± 1,8E+04	± 2,3E+04	± 1,8E+04
	0,366		0,280	0,722	0,827
<i>Ratio</i>	2,5		1,9	2,5	2,6
<i>Lymphocytes T</i>	± 0,2		± 0,2	± 0,2	± 0,2
<i>CD4+ et</i>	1,8		1,8	1,8	2,1
<i>CD8+</i>	± 0,2		± 0,2	± 0,2	± 0,2
<i>0,063</i>	0,041		0,484	0,027	0,116
<i>IL1a</i>	30,6	8,5	29,4	27,2	26,7
<i>0,759</i>	± 8,3	± 2,5	± 8,0	± 7,4	± 7,3
	31,8	12,8	26,8	28,1	29,2
	± 8,6	± 3,6	± 7,3	± 7,7	± 8,0
	0,922	0,355	0,817	0,939	0,822
					0,640
					0,713

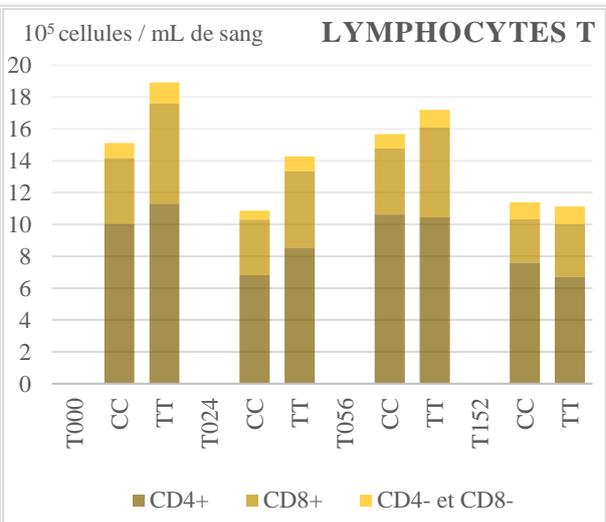
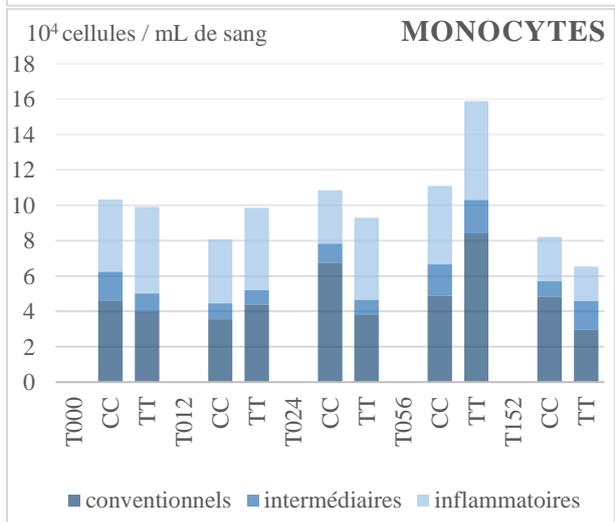
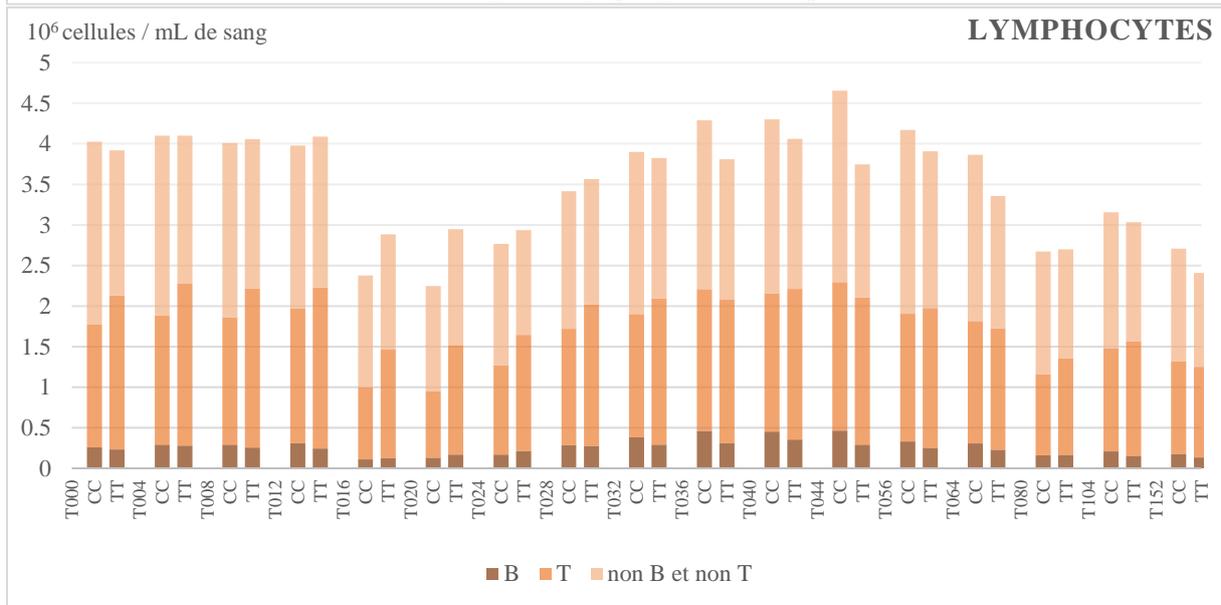
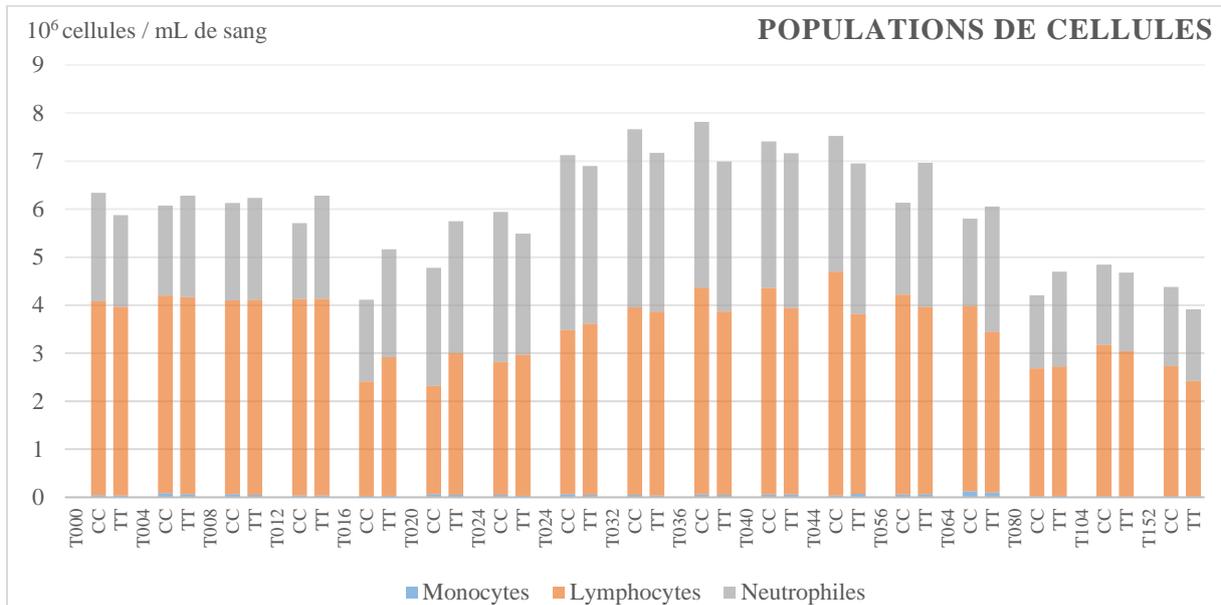
<i>IL1b</i>	12,5	12,6	10,7	13,9	13,4	11,0	12,7
<i>0,138</i>	± 4,5	± 4,5	± 3,9	± 4,9	± 4,8	± 4,0	± 4,6
	26,1	25,0	18,8	26,8	30,0	27,4	24,1
	± 9,0	± 8,6	± 6,6	± 9,2	± 10,3	± 9,4	± 8,3
	0,200	0,226	0,311	0,241	0,169	0,135	0,254
<i>IL1Ra</i>	139,4	30,4	102,0	100,3	109,3	86,3	130,6
<i>0,955</i>	± 67,8	± 15,2	± 49,8	± 49,0	± 53,3	± 42,2	± 63,6
	121,0	38,8	101,2	115,8	104,4	108,6	111,8
	± 58,9	± 19,2	± 49,4	± 56,4	± 50,9	± 53,0	± 54,5
	0,841	0,737	0,991	0,839	0,948	0,747	0,826
<i>IL2</i>	1214,0	427,9	1141,3	1001,4	1016,7	1040,9	896,7
<i>0,932</i>	± 400,6	± 141,4	± 376,6	± 330,5	± 335,5	± 343,5	± 296,0
	1279,6	611,5	1129,1	1187,0	1174,2	511,7	1100,1
	± 422,2	± 201,9	± 372,6	± 391,7	± 387,5	± 169,1	± 363,1
	0,912	0,471	0,982	0,724	0,764	0,192	0,672
<i>IL4</i>	983,0	547,2	931,8	904,4	990,1	924,6	966,6
<i>0,643</i>	± 257,6	± 143,5	± 244,2	± 237,0	± 259,4	± 242,3	± 253,3
	1146,0	800,6	1025,4	1110,0	1140,6	1125,9	989,4
	± 300,2	± 209,8	± 268,7	± 290,8	± 298,8	± 295,0	± 259,3
	0,688	0,339	0,801	0,594	0,710	0,608	0,951
<i>IL6</i>	16,4	6,0	23,4	348,9	224,1	16,9	166,1
<i>0,763</i>	± 9,4	± 3,8	± 13,1	± 188,8	± 121,4	± 9,6	± 90,2
	18,6	6,1	30,9	328,3	131,5	68,7	175,4
	± 10,6	± 3,9	± 17,2	± 177,7	± 71,5	± 37,6	± 95,2
	0,877	0,979	0,733	0,938	0,524	0,207	0,945
<i>IL8</i>	1099,9	1066,2	1362,7	1289,9	1325,2	1489,2	1577,9
<i>0,093</i>	± 139,7	± 135,4	± 173,0	± 163,8	± 168,3	± 189,1	± 200,3
	1729,2	1388,2	1654,5	1529,1	1473,8	1792,3	2076,9
	± 219,5	± 176,3	± 210,0	± 194,1	± 187,1	± 227,5	± 263,6
	0,032	0,173	0,305	0,365	0,566	0,326	0,158

<i>IL10</i> 0,692	554,4	196,7	567,9	823,2	567,1	487,1	509,1
	± 143,4	± 51,1	± 146,9	± 212,9	± 146,7	± 126,1	± 131,7
	474,1	213,8	430,5	659,9	491,4	503,4	386,2
	± 122,7	± 55,5	± 111,4	± 170,7	± 127,2	± 130,3	± 100,0
	0,678	0,825	0,471	0,561	0,703	0,930	0,472
<i>IL17a</i> 0,602	2,0	1,1	2,8	5,7	2,3	2,1	4,4
	± 0,9	± 0,7	± 1,2	± 2,1	± 1,0	± 1,0	± 1,7
	1,7	0,8	1,8	3,1	2,1	2,1	2,2
	± 0,9	± 0,6	± 0,9	± 1,3	± 1,0	± 1,0	± 1,0
	0,826	0,756	0,524	0,328	0,919	0,976	0,303
<i>IFNg</i> 0,890	2,3	1,6	2,2	5,3	10,8	2,8	1,9
	± 0,9	± 0,7	± 0,8	± 1,7	± 3,1	± 1,0	± 0,8
	2,7	2,0	2,7	5,2	9,4	2,9	2,2
	± 1,0	± 0,8	± 1,0	± 1,6	± 2,7	± 1,0	± 0,8
	0,786	0,747	0,685	0,943	0,741	0,976	0,816
<i>IP10</i> 0,822	1088,0	671,7	916,2	911,3	1426,8	982,6	537,5
	± 178,9	± 110,5	± 150,7	± 149,9	± 234,6	± 161,6	± 88,5
	1245,8	795,6	972,7	934,1	1205,0	934,3	607,9
	± 204,8	± 130,9	± 160,0	± 153,6	± 198,1	± 153,7	± 100,0
	0,573	0,483	0,802	0,917	0,484	0,832	0,608
<i>CCL2</i> 0,155	518,9	273,0	695,3	1843,1	881,1	548,6	932,3
	± 107,8	± 56,8	± 144,4	± 382,3	± 182,9	± 113,9	± 193,5
	434,2	242,1	469,5	957,2	551,3	441,4	561,6
	± 90,2	± 50,4	± 97,5	± 198,7	± 114,5	± 91,7	± 116,7
	0,558	0,691	0,219	0,062	0,152	0,478	0,127
<i>CCL3</i> 0,198	642,6	280,5	631,4	613,1	608,6	634,0	613,9
	± 81,2	± 35,5	± 79,8	± 77,4	± 76,9	± 80,1	± 77,5
	803,6	405,9	737,7	780,2	762,4	793,7	698,6
	± 101,5	± 51,3	± 93,2	± 98,5	± 96,3	± 100,2	± 88,2
	0,239	0,067	0,403	0,207	0,236	0,237	0,484

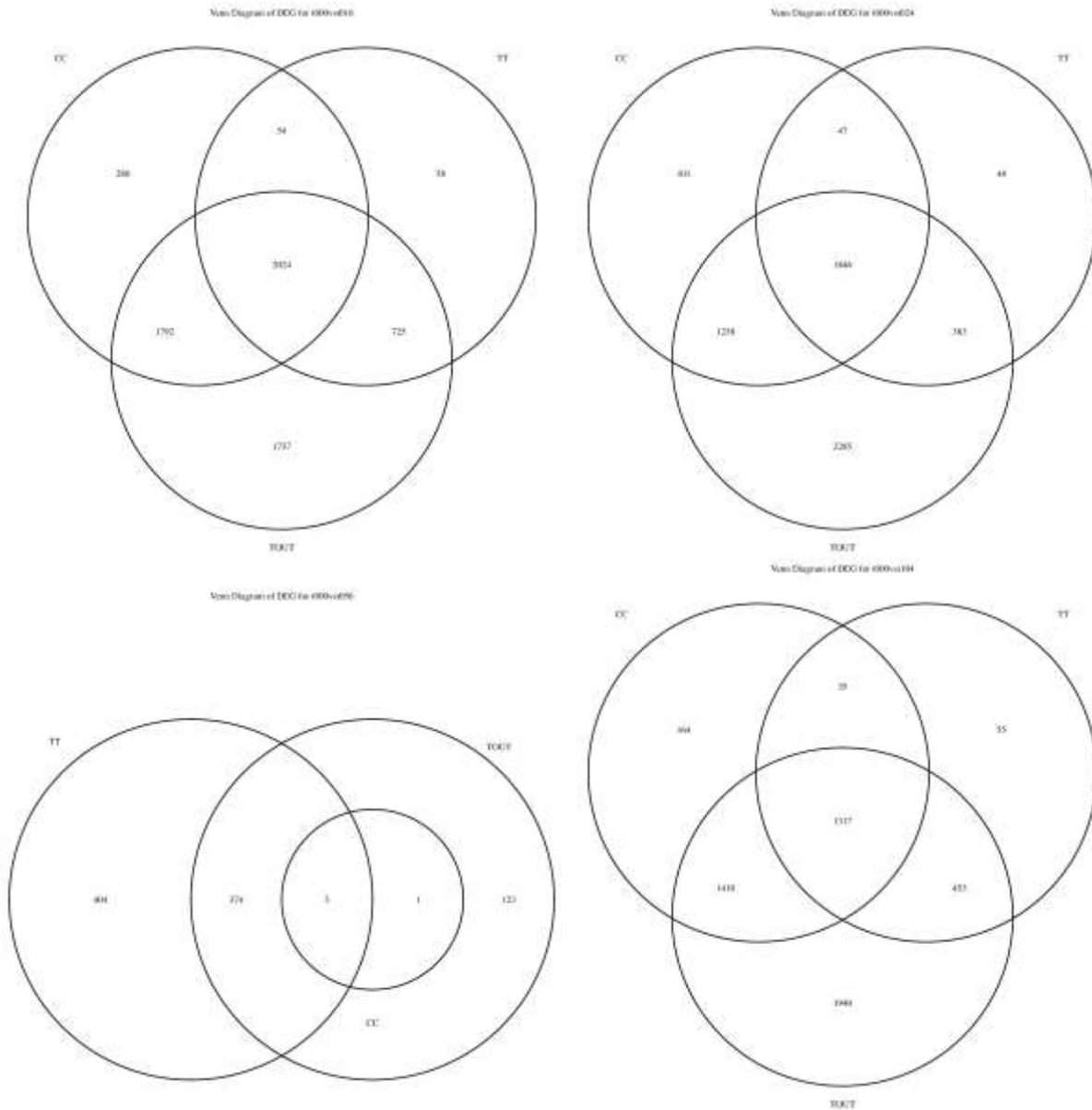
<i>CCL4</i>	40,4	7,7	37,5	39,0	39,3	37,5	36,3
	± 13,3	± 2,8	± 12,4	± 12,8	± 12,9	± 12,4	± 12,0
<i>0,741</i>	1255,1	608,1	1112,3	1281,2	1182,6	1326,6	921,8
	± 9,5	± 3,4	± 9,0	± 12,5	± 10,9	± 12,6	± 8,7
	0,482	0,682	0,510	0,949	0,715	0,964	0,504
<i>TNFα</i>	1244,5	418,6	1194,2	1092,6	1164,0	1129,1	923,2
	± 370,6	± 124,9	± 355,7	± 325,4	± 346,7	± 336,3	± 275,0
<i>0,822</i>	1255,1	608,1	1112,3	1281,2	1182,6	1326,6	921,8
	± 373,8	± 181,2	± 331,3	± 381,6	± 352,2	± 395,1	± 274,6
	0,984	0,406	0,869	0,714	0,971	0,710	0,997

Abréviation : inoc. : inoculée. Les p-valeurs en gras et surlignées en jaune montrent un effet significatif de SOCS2 sur le phénotype étudié (p-valeur < 0,05) tandis que les p-valeurs surlignées en vert sont celles qui se rapprochent le plus de la significativité (0,05 ≤ p-valeur < 0,1) et pour lesquelles une interprétation de tendance est réalisée au vu du faible effectif de l'expérimentation (seuil suggestif).

Annexe 11 Quantités moyennes par génotype SOCS2 des populations cellulaires dans le sang au cours de l'infection expérimentale (Chapitre IV.)



Annexe 12 Diagrammes de Venn représentant les gènes différentiels en commun ou indépendants entre les analyses RNA-seq mettant en lumière la réponse immunitaire (différences entre les temps, analyses tous génotypes SOCS2 -TOUT- ou intra-génotype SOCS2 -CC et TT-) entre T0 et les autres temps (quand le nombre de gènes est différent de 0 ou 1) ([Chapitre IV.](#))



Annexe 13 Principales (Top 10) voies canoniques parmi les gènes à expression différentielle entre quelques temps de l'infection expérimentale ([Chapitre IV.](#))

Contraste	Gènes annotés	Voies canoniques
<i>T0 vs T16</i>	3 940 (63%)	<ol style="list-style-type: none"> 1. ERK/MAPK 2. Aldostérone dans les cellules épithéliales 3. Activation de Th1 and Th2 4. Mécanismes moléculaires du cancer 5. PI3K/AKT 6. PI3K dans les lymphocytes B 7. Récepteur des lymphocytes B 8. CD28 dans les lymphocytes T auxiliaires 9. Th1 10. Récepteur des lymphocytes T
<i>T0 vs T24</i>	3 919 (68%)	<ol style="list-style-type: none"> 1. Mécanismes moléculaires du cancer 2. CD28 dans les lymphocytes T auxiliaires 3. Rôle du gène <i>BRCA1</i> dans la réponse en cas de dégradation de l'ADN 4. Récepteur des lymphocytes T 5. Récepteur des lymphocytes B 6. Cancer du sein génétique 7. Activation de Th1 and Th2 8. mTOR 9. Phagocytose via les récepteurs FCγ dans les macrophages et monocytes 10. RANK dans les ostéoclastes
<i>T0 vs T56</i>	277 (55%)	<ol style="list-style-type: none"> 1. Rôles mitotiques de la PLK 2. Contrôle du cycle cellulaire de la réplication chromosomique 3. Entrée en phase S à médiation œstrogénique 4. Ostéoarthrose 5. Cycle cellulaire : Régulation des points de contrôle G2/M des dégradations de l'ADN 6. Interférons 7. Inhibition des métalloprotéases matricielles 8. GADD45 9. Dégradation de l'ADN induite par 14-3-3σ 10. Cyclines et régulation du cycle cellulaire
<i>T0 vs T104</i>	3 371 (66%)	<ol style="list-style-type: none"> 1. Hypertrophie cardiaque (augmentée) 2. Intégrines 3. Phagocytose via les récepteurs FCγ dans les macrophages et monocytes 4. Extravasation des leukocytes 5. STAT3 6. Activation de Th1 and Th2 7. Mécanismes moléculaires du cancer 8. Jonction des cellules germinales et de Sertoli 9. Kinases Tec 10. Récepteur des lymphocytes T

Abréviations : *BRCA1* : *BReast CAncer 1* ; *ERK* : *Extracellular-signal-Regulated Kinase* ; *FC* : *Fragment cristallisable* ; *GADD45* : *Growth Arrest and Dna Damage inducible* ; *MAPK* : *Mitogen-Activated Protein Kinases* ; *PI3K* : *PhosphoInositide 3-Kinase* ; *PLK* : *Polo-Like Kinase* ; *RANK* : *Receptor Activator of Nuclear Factor kappa-B* ; *STAT* : *Signal Transducers and Activators of Transcription* ; *Th* : *Lymphocytes T auxiliaires*.

Annexe 14 Gènes différentiellement exprimés entre les C/C et les T/T aux différents temps de l'infection expérimentale ([Chapitre IV.](#))

<i>P-valeur</i>	<i>Gène différentiellement exprimé</i>	<i>Log fold-change</i>	<i>OAR</i>	<i>Début (Mb, sauf pour les Contig, en pb)</i>	<i>Fin</i>	<i>Temps</i>
1,75E-09	<i>LOC105606568</i>	2,92	4	70,55	70,56	t056
1,03E-08	<i>XLOC_000330</i>	-5,64	Contig1734	16 705	18 048	t104
6,61E-08	<i>XLOC_005504</i>	4,50	1	292,96	292,97	t056
2,13E-07	<i>PTX3</i>	6,04		250,84	250,85	t056
2,96E-07	<i>LOC114108696</i>	-4,22	2	217,07	217,08	t008
3,22E-07	<i>CMPK2</i>	1,63	3	17,27	17,29	t056
3,44E-07	<i>LOC105606856</i>	1,59	11	48,92	48,93	t056
4,02E-07	<i>SMTNL1/UBE2L6</i>	1,17	15	87,01	87,12	t056
8,91E-07	<i>IFI6</i>	2,48	2	255,15	255,18	t056
1,00E-06	<i>IL3RA</i>	2,61	X	0,78	0,78	t056
1,06E-06	<i>FAM71F1</i>	2,17	4	101,51	101,55	t056
1,08E-06	<i>XLOC_029031</i>	3,98	3	160,91	160,94	t056
1,29E-06	<i>XLOC_024007</i>	2,32	21	30,70	30,71	t056
1,34E-06	<i>SLC26A8</i>	2,29	20	11,56	11,64	t056
1,48E-06	<i>XLOC_039743</i>	2,69	9	54,60	54,60	t056
1,53E-06	<i>XLOC_000627</i>	-5,50	Contig2315	4 141	5 143	t056
1,70E-06	<i>SDR42E2</i>	-4,65	24	20,54	20,56	t056
2,37E-06	<i>XLOC_000401</i>	-2,51	Contig1890	11 070	14 207	t012
2,43E-06	<i>XLOC_010460</i>	2,89	13	49,75	49,77	t016
2,63E-06	<i>IQCH</i>	4,24	7	15,02	15,26	t056
2,64E-06	<i>LOC114110327</i>	6,49	22	43,42	43,42	t056
3,07E-06	<i>LOC114116917</i>	4,31	11	49,60	49,61	t056
3,17E-06	<i>LOC101103862 LOC101104114</i>	4,43	23	68,06	68,11	t056
3,22E-06	<i>NR2E1</i>	2,63	8	32,29	32,29	t056
3,46E-06	<i>MINDY1</i>	1,84	1	106,69	106,70	t056
3,86E-06	<i>BMX</i>	2,52	X	14,38	14,43	t056
4,14E-06	<i>DPH5</i>	1,47	1	82,85	82,98	t056
4,48E-06	<i>BRSK2</i>	-1,92	21	51,94	51,99	t056
4,63E-06	<i>LOC114110327</i>	6,26	22	43,42	43,42	t104
5,17E-06	<i>XLOC_024327</i>	2,93	21	52,69	52,69	t056
5,21E-06	<i>LOC101107096</i>	3,08	5	16,23	16,32	t056
5,59E-06	<i>LOC114118421</i>	3,64	15	51,35	51,36	t104
6,11E-06	<i>IL22RA2</i>	2,36	8	68,45	68,55	t056
6,86E-06	<i>S100A9</i>	3,06	1	109,83	109,83	t056
6,88E-06	<i>XLOC_034619</i>	5,29	5	16,23	16,23	t056
6,88E-06	<i>LOC114118421</i>	3,55	15	51,35	51,36	t056
6,92E-06	<i>LOC114113694</i>	1,93	3	16,04	16,13	t056

7,31E-06	<i>XLOC_032661</i>	2,87	4	34,82	34,83	t056
7,97E-06	<i>XLOC_000627</i>	-5,26	Contig2315	4 141	5 143	t104
7,98E-06	<i>FZD5</i>	1,98	2	223,94	223,96	t056
8,51E-06	<i>LOC105607012</i>	2,43		27,68	27,70	t056
8,62E-06	<i>ALPL</i>	2,74		260,71	260,78	t056
8,94E-06	<i>DDX58</i>	2,32	2	109,83	109,89	t056
9,41E-06	<i>FZD6</i>	2,42	9	81,99	82,03	t104
9,69E-06	<i>AP1M2</i>	2,29	5	13,40	13,41	t104
9,85E-06	<i>LOC114114833</i>	2,96		15,81	15,84	t056
9,98E-06	<i>S100A8</i>	2,91	1	109,86	109,87	t056
1,05E-05	<i>ADAMTSL5</i>	-3,34	5	46,01	46,02	t056
1,13E-05	<i>RSAD2</i>	2,00	3	17,30	17,31	t056
1,18E-05	<i>LOC114113097/XLOC_003973</i>	3,45	1	75,97	76,06	t056
1,26E-05	<i>TCN1</i>	2,92	15	89,36	89,38	t056
1,34E-05	<i>RAB3IP</i>	2,07	3	160,84	160,91	t056
1,47E-05	<i>XLOC_039596</i>	2,31	9	25,41	25,41	t056
1,51E-05	<i>IL1R2</i>	3,85	3	106,08	106,25	t056
1,69E-05	<i>GPAT2</i>	5,70		110,50	110,51	t104
1,73E-05	<i>XLOC_027597</i>	4,16	26	43,54	43,54	t056
1,77E-05	<i>LOC105611303</i>	1,87	1	169,80	169,83	t056
1,82E-05	<i>XLOC_035000</i>	2,11	5	46,45	46,46	t056
2,06E-05	<i>XLOC_037592</i>	2,12	7	18,69	18,74	t056
2,09E-05	<i>CRYAB</i>	3,44	15	24,75	24,77	t056
2,16E-05	<i>LOC105605445/MIR221</i>	1,30	X	46,23	46,34	t056
2,18E-05	<i>EIF2AK2</i>	1,22	3	92,51	92,55	t056
2,36E-05	<i>XLOC_029029</i>	4,16		160,91	160,91	t056
2,38E-05	<i>ADGRG3</i>	2,20	14	27,25	27,28	t056
2,40E-05	<i>FOLR3</i>	4,58	15	54,79	54,80	t056
2,41E-05	<i>LOC114108712</i>	2,36	17	70,45	70,46	t056
2,53E-05	<i>ASAP1/XLOC_040120</i>	2,82	9	25,07	25,41	t056
2,54E-05	<i>BCL2A1/LOC101114310/XLOC_016610</i>	1,49	18	23,85	23,92	t056
2,62E-05	<i>XLOC_012598</i>	2,29	14	61,18	61,19	t056
2,87E-05	<i>IL18RAP</i>	2,06	3	105,69	105,73	t056
2,91E-05	<i>IL18R1</i>	2,07		105,74	105,77	t056
2,97E-05	<i>LOC114112479/XLOC_000758</i>	-1,67	Contig267	10 007	23 706	t056
3,21E-05	<i>RYR1</i>	2,17	14	51,12	51,25	t056
3,37E-05	<i>LOC101103771</i>	2,55	1	109,84	109,85	t056
3,51E-05	<i>XLOC_033332/ZNF333</i>	2,08	5	9,35	9,44	t056
3,56E-05	<i>LOC106990969</i>	1,97	2	223,84	223,89	t056
3,66E-05	<i>IP6K3</i>	6,80	20	9,55	9,57	t056
3,83E-05	<i>BST-2A</i>	2,01	5	5,75	5,76	t056
3,84E-05	<i>AUH/XLOC_020296</i>	1,45	2	27,72	27,92	t056

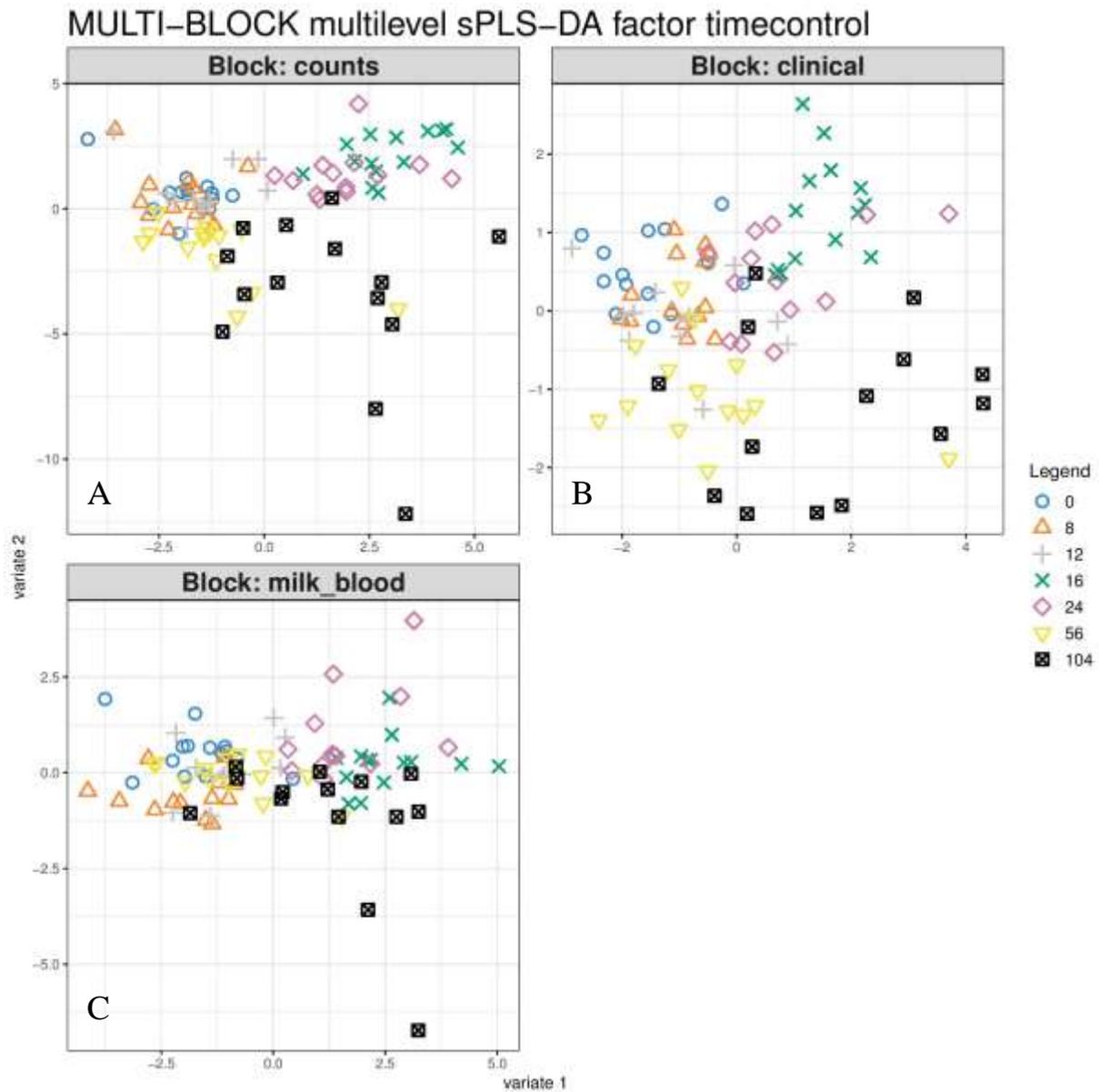
3,94E-05	<i>XLOC_037579</i>	3,62	7	17,34	17,35	t056
3,99E-05	<i>LOC101112257</i>	3,04	15	89,41	89,42	t056
4,04E-05	<i>XLOC_015955</i>	5,92	17	23,88	23,88	t056
4,27E-05	<i>THBD</i>	2,10	13	43,27	43,28	t056
4,33E-05	<i>LOC101102413</i>	4,73	14	41,86	41,86	t056
4,34E-05	<i>LOC101102627</i>	2,14	2	142,57	142,57	t056
4,71E-05	<i>SETD1B</i>	5,83	17	62,45	62,48	t056
4,82E-05	<i>XLOC_042226</i>	1,25	X	115,66	115,67	t056
4,84E-05	<i>CFI</i>	2,98	6	18,73	18,78	t056
4,99E-05	<i>LOC101110828</i>	-2,46	14	63,77	63,95	t056
5,11E-05	<i>CHMP5</i>	0,61	2	39,90	39,91	t056
5,17E-05	<i>MMP9</i>	2,97	13	78,11	78,12	t056
5,30E-05	<i>LOC105608615 LOC114113997</i>	-1,66	3	155,32	155,48	t056
5,56E-05	<i>XLOC_023085</i>	5,88	20	41,76	41,76	t056
5,62E-05	<i>XLOC_020177</i>	3,38	2	3,21	3,23	t056
5,86E-05	<i>FABP7.3</i>	1,88	8	17,29	17,29	t056
5,88E-05	<i>MX2</i>	3,07	1	284,81	284,87	t056
5,99E-05	<i>CASP7</i>	0,81	22	37,31	37,35	t056
6,05E-05	<i>XLOC_035548</i>	6,58	6	46,58	46,59	t056
6,26E-05	<i>XLOC_005810</i>	5,15	10	57,34	57,40	t056
6,96E-05	<i>RTP4</i>	1,48	1	218,44	218,44	t056
7,01E-05	<i>GPAT3</i>	1,31	6	108,60	108,66	t056
7,16E-05	<i>FAR2</i>	1,55	3	198,76	198,93	t056
7,17E-05	<i>LOC101103965</i>	1,37		194,02	194,06	t056
7,22E-05	<i>XLOC_016625</i>	3,50	18	25,23	25,24	t056
7,32E-05	<i>XLOC_022290</i>	3,02	20	40,78	40,78	t056
7,77E-05	<i>XLOC_029779</i>	-3,21	3	238,20	238,21	t056
7,88E-05	<i>XLOC_035910</i>	2,55	6	123,05	123,06	t056
8,12E-05	<i>LOC101107365</i>	1,46	10	13,57	13,58	t056
8,18E-05	<i>IFIT3</i>	1,64	22	12,76	12,77	t056
8,44E-05	<i>XLOC_036723</i>	-3,49	7	12,35	12,35	t056
8,63E-05	<i>IFI35</i>	0,60	11	20,07	20,08	t056
8,81E-05	<i>IL2RA</i>	1,45	13	11,38	11,43	t056
8,92E-05	<i>LOC114118420 XLOC_013297</i>	1,32	15	51,16	51,25	t056
9,17E-05	<i>LOC105606875</i>	7,22	11	49,55	49,59	t056
9,23E-05	<i>MGAM</i>	2,16	4	114,11	114,20	t056
9,27E-05	<i>XLOC_024009</i>	2,37	21	30,70	30,70	t056
9,66E-05	<i>BEST3</i>	3,88	3	160,96	161,02	t056
1,04E-04	<i>LOC105610700</i>	3,58	4	34,86	34,91	t056
1,05E-04	<i>VEPH1</i>	3,22	1	250,76	251,03	t056
1,06E-04	<i>XLOC_039754</i>	4,55	9	54,84	54,84	t056
1,08E-04	<i>SERINC2</i>	1,65	2	251,61	251,63	t056

1,10E-04	<i>LOC105608648</i>	4,90	13	83,36	83,38	t056
1,11E-04	<i>TNFSF13B</i>	1,35	10	93,73	93,76	t056
1,12E-04	<i>LOC101104700</i>	1,75	21	52,70	52,70	t056
1,14E-04	<i>ANXA3</i>	1,77	6	103,30	103,38	t056
1,19E-04	<i>LMNB1</i>	0,88	5	25,75	25,81	t056
1,20E-04	<i>LOC114118454</i>	-3,55	15	72,23	72,23	t056
1,22E-04	<i>FAM129A</i>	1,48	12	70,50	70,67	t056
1,25E-04	<i>XLOC_026215</i>	2,28	24	41,14	41,21	t056
1,28E-04	<i>LOC114110138</i>	1,17	21	52,68	52,68	t056
1,35E-04	<i>TNFAIP8</i>	0,72	5	34,17	34,31	t056
1,38E-04	<i>TRNAV-CAC</i>	1,60	1	105,17	105,17	t056
1,39E-04	<i>NT5C3A</i>	1,21	4	70,53	70,61	t056
1,46E-04	<i>XLOC_000906</i>	-1,96	Contig488	25 515	28 479	t056
1,52E-04	<i>XLOC_024102</i>	2,59	21	41,97	42,00	t056
1,56E-04	<i>IFIH1</i>	0,91	2	159,36	159,42	t056
1,56E-04	<i>LOC114118716</i>	3,02	16	2,65	2,68	t056
1,69E-04	<i>XLOC_031755</i>	-1,23	3	240,19	240,20	t056
1,71E-04	<i>RALB</i>	1,01	2	198,90	199,01	t056
1,78E-04	<i>XLOC_018015</i>	1,79	19	55,22	55,22	t056
1,92E-04	<i>LOC105603222</i>	1,05	18	58,00	58,01	t056
2,04E-04	<i>XLOC_031753</i>	-1,12	3	240,18	240,20	t056
2,04E-04	<i>PARM1</i>	1,21	6	99,72	99,85	t056
2,14E-04	<i>MEGF9</i>	1,54	2	2,87	2,95	t056
2,15E-04	<i>DCDC2</i>	4,74	20	35,91	36,06	t056
2,22E-04	<i>XLOC_015961</i>	2,95	17	28,56	28,57	t056
2,35E-04	<i>LOC101104797</i>	2,26	1	247,12	247,13	t056
2,40E-04	<i>GYG1</i>	1,08		260,41	260,45	t056
2,41E-04	<i>DHX58</i>	1,14	11	20,89	20,90	t056
2,48E-04	<i>TNFAIP6 XLOC_021074</i>	2,09	2	169,84	169,90	t056
2,51E-04	<i>UPPI</i>	1,80	4	9,10	9,17	t056
2,60E-04	<i>FAM217B</i>	1,13	13	58,96	58,96	t056
2,69E-04	<i>LOC114113838</i>	-1,87	3	241,01	241,01	t056
2,69E-04	<i>FAM32A</i>	0,53	5	6,98	6,99	t056
2,70E-04	<i>LOC101103588</i>	1,29	14	61,66	61,67	t056
2,73E-04	<i>LOC105608827</i>	2,97		50,49	50,50	t056
2,79E-04	<i>SLCO4A1</i>	1,73	13	57,03	57,05	t056
2,86E-04	<i>LOC101121908</i>	2,28	1	109,79	109,80	t056
2,88E-04	<i>XLOC_008939</i>	-2,58	12	4,03	4,03	t056
2,89E-04	<i>WDR78</i>	1,80	1	44,68	44,79	t056
3,03E-04	<i>MX1</i>	1,29	1	284,88	284,88	t056
3,09E-04	<i>TDRD7</i>	0,77	2	52,82	52,96	t056
3,12E-04	<i>LOC105613298 LOC106990575 LOC114108669 LOC114117851</i>	-1,54	14	7,95	8,06	t056

3,14E-04	<i>XLOC_002900</i>	3,25	1	217,87	217,88	t056
3,29E-04	<i>XLOC_005813</i>	4,16	10	57,48	57,48	t056
3,34E-04	<i>XLOC_006140</i>	3,46		32,50	32,51	t056
3,36E-04	<i>ONECUT2</i>	4,11	23	62,77	62,83	t056
3,54E-04	<i>NFIL3</i>	1,34	2	27,97	27,99	t056
3,55E-04	<i>LOC105602034</i>	0,47	14	70,17	70,18	t056
3,67E-04	<i>LOC114111192</i>	3,03	1	82,81	82,85	t056
3,67E-04	<i>XLOC_024008</i>	1,94	21	30,70	30,70	t056
3,74E-04	<i>LOC114113969</i>	2,75	3	141,75	141,75	t056
3,74E-04	<i>LOC101105064 PNP</i>	1,26	5	43,67	43,69	t056
3,78E-04	<i>LOC105608949</i>	2,47	2	141,28	141,33	t056
3,83E-04	<i>TFF2</i>	2,10	1	285,93	285,93	t056
3,89E-04	<i>CMTR2</i>	0,76	14	42,57	42,58	t056
4,04E-04	<i>LOC114112166</i>	-2,38	Contig1696	14 297	15 282	t056
4,12E-04	<i>SLC25A38</i>	0,46	19	13,23	13,25	t056
4,13E-04	<i>GADD45A</i>	1,32	1	45,51	45,52	t056
4,17E-04	<i>LOC114113739</i>	0,95	3	115,96	116,00	t056
4,20E-04	<i>LCN2</i>	4,71		8,37	8,37	t056
4,26E-04	<i>LOC114109133</i>	-1,19	18	63,36	63,42	t056
4,27E-04	<i>LOC114114843</i>	1,41	5	18,90	18,90	t056
4,28E-04	<i>E2F2</i>	0,83	2	258,85	258,88	t056
4,28E-04	<i>CSF2RB</i>	1,13	3	193,38	193,40	t056
4,32E-04	<i>XLOC_031757</i>	-1,55		240,20	240,21	t056
4,43E-04	<i>IGSF6</i>	1,64	24	20,82	20,83	t056

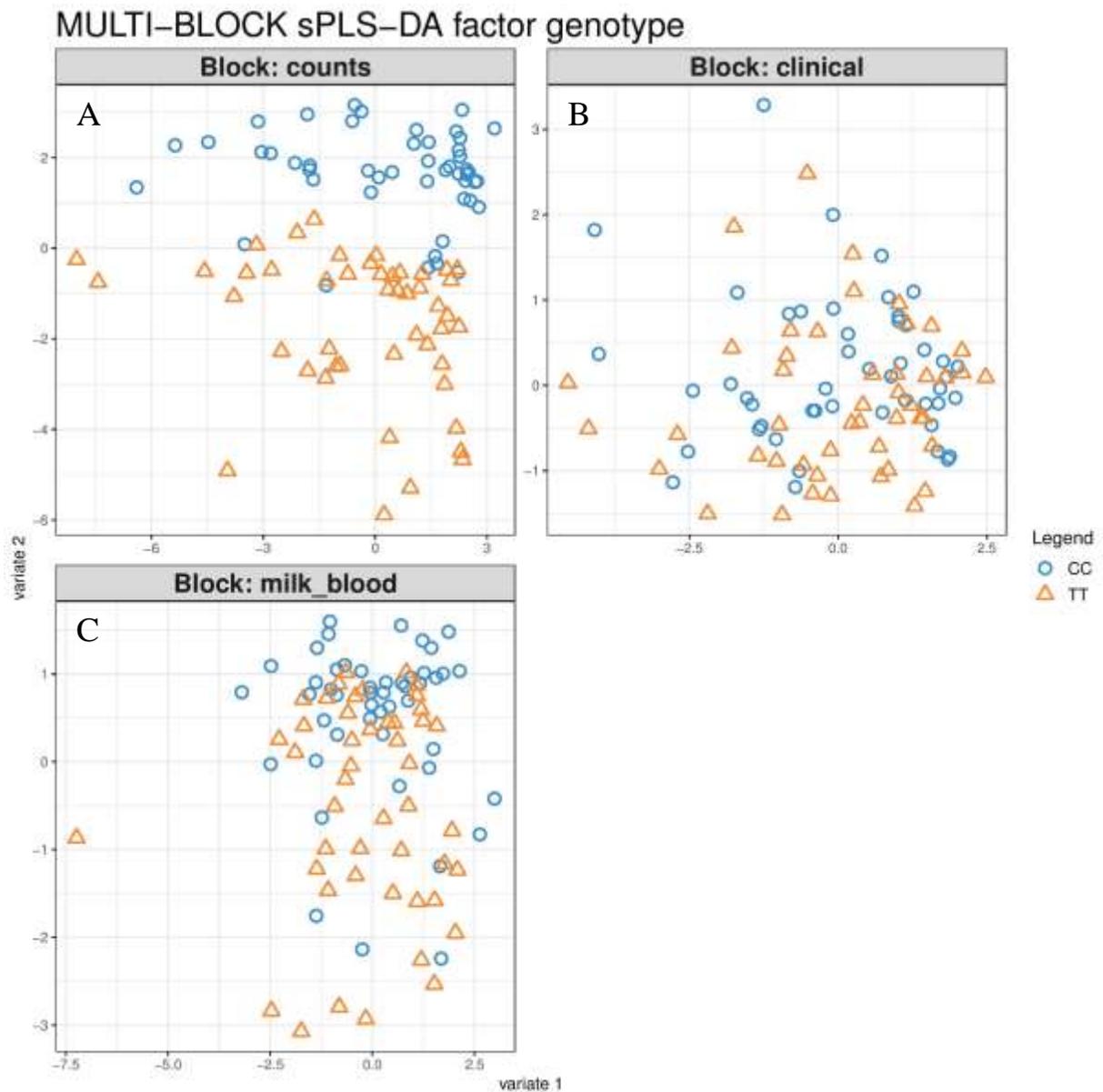
Abréviations : OAR : Ovis Aries ; Mb : Mégabase ; pb : paire de bases. Le fold-change (exprimé en logarithme en base 2) est le rapport du niveau moyen d'expression du gène chez les T/T par rapport à la référence C/C. Les gènes commençant par « LOC » sont des loci non annotés spécifiquement pour l'espèce ovine sur le génome de référence. Ceux commençant par « XLOC » sont des loci nouvellement créés lors de l'assemblage de novo sur notre nouveau modèle de référence à partir des transcrits de cette infection expérimentale. Lors de l'assemblage de novo, des loci qui englobent plusieurs loci peuvent apparaître : ils proviennent de lectures qui font la jonction entre plusieurs gènes car la polymérase ne s'arrête pas.

Annexe 15 Sparse multilevel multi-bloc PLS-DA avec comme facteur discriminant le « temps de contrôle » sur les blocs des données de comptages de gènes (A), des données cliniques (B), et des données lait-sang (C) à tous les temps du transcriptome de l'infection expérimentale ([Chapitre IV.](#))



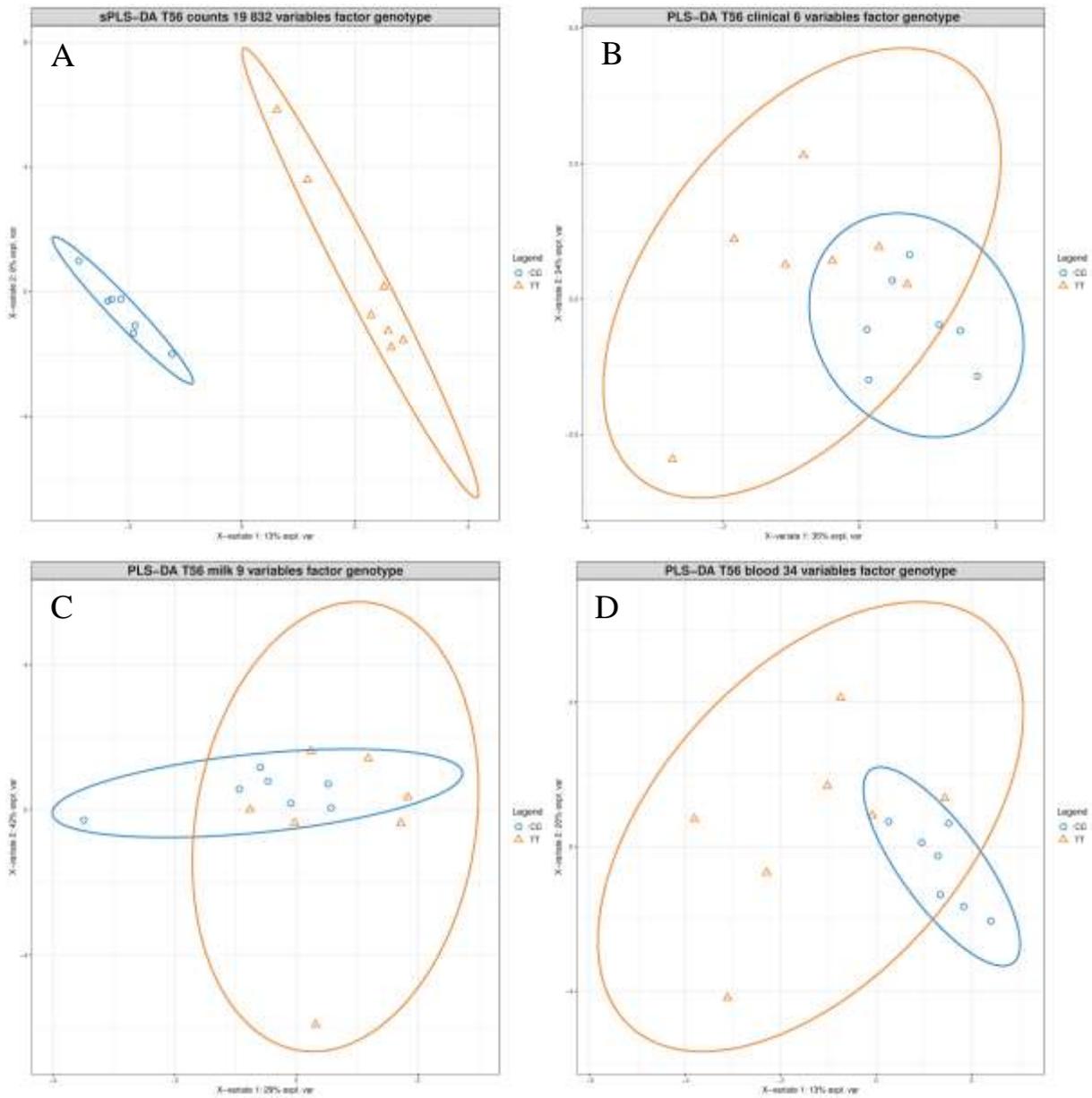
Chaque temps de contrôle est matérialisé par une couleur et un symbole différents, comme indiqué dans la légende. Chaque point symbolisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole.

Annexe 16 Sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » sur les blocs des données de comptages de gènes (A), des données cliniques (B), et des données lait-sang (C) à tous les temps du transcriptome de l'infection expérimentale ([Chapitre IV.](#))



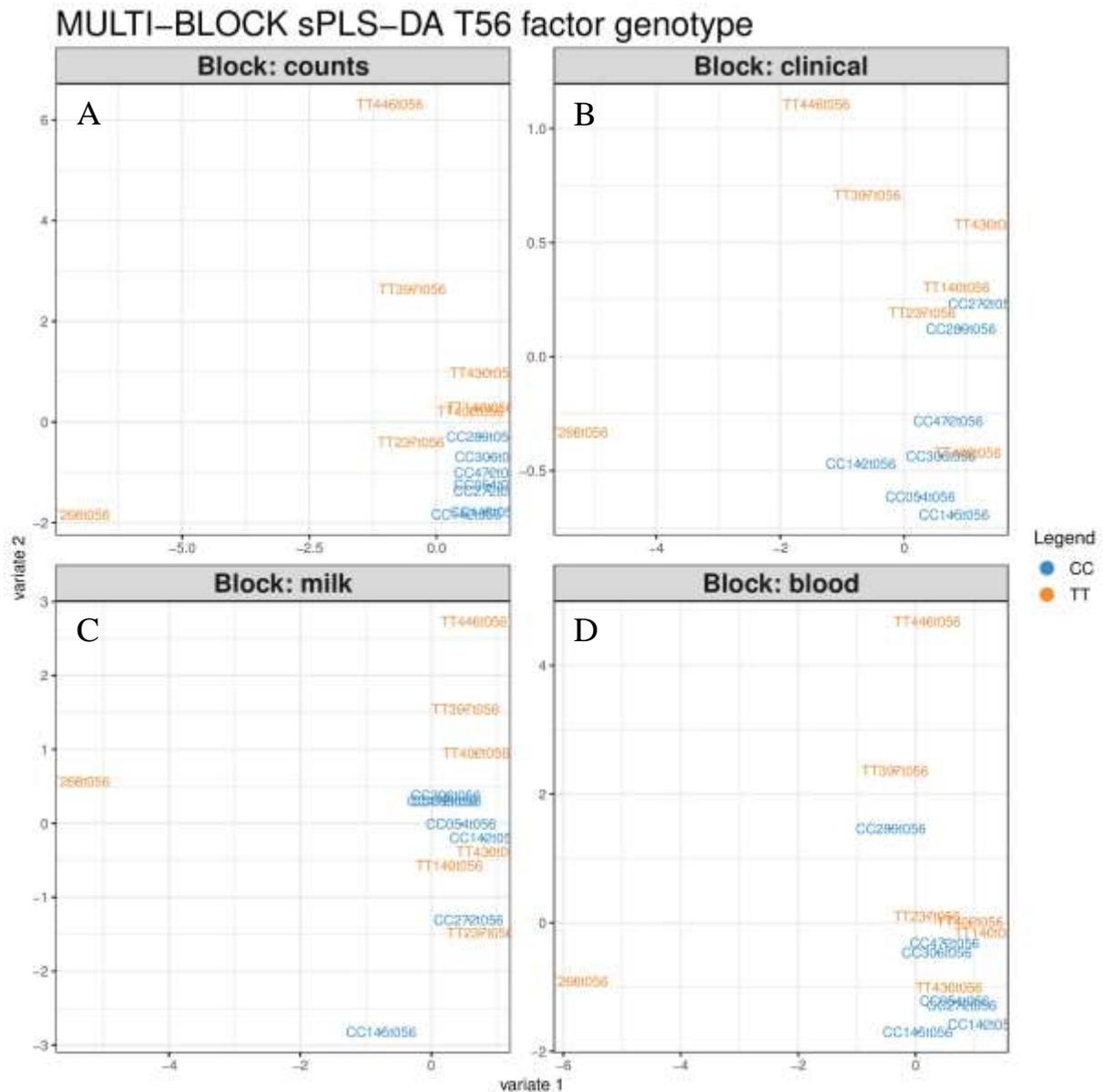
Chaque génotype SOCS2 est matérialisé par une couleur et un symbole différents, comme indiqué dans la légende. Chaque point symbolisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole.

Annexe 17 PLS-DA avec comme facteur discriminant le « génotype SOCS2 » sur les blocs des données de comptages de gènes (A, avec option « sparse »), cliniques (B), de lait (C), et de sang (D) à T56 ([Chapitre IV.](#))



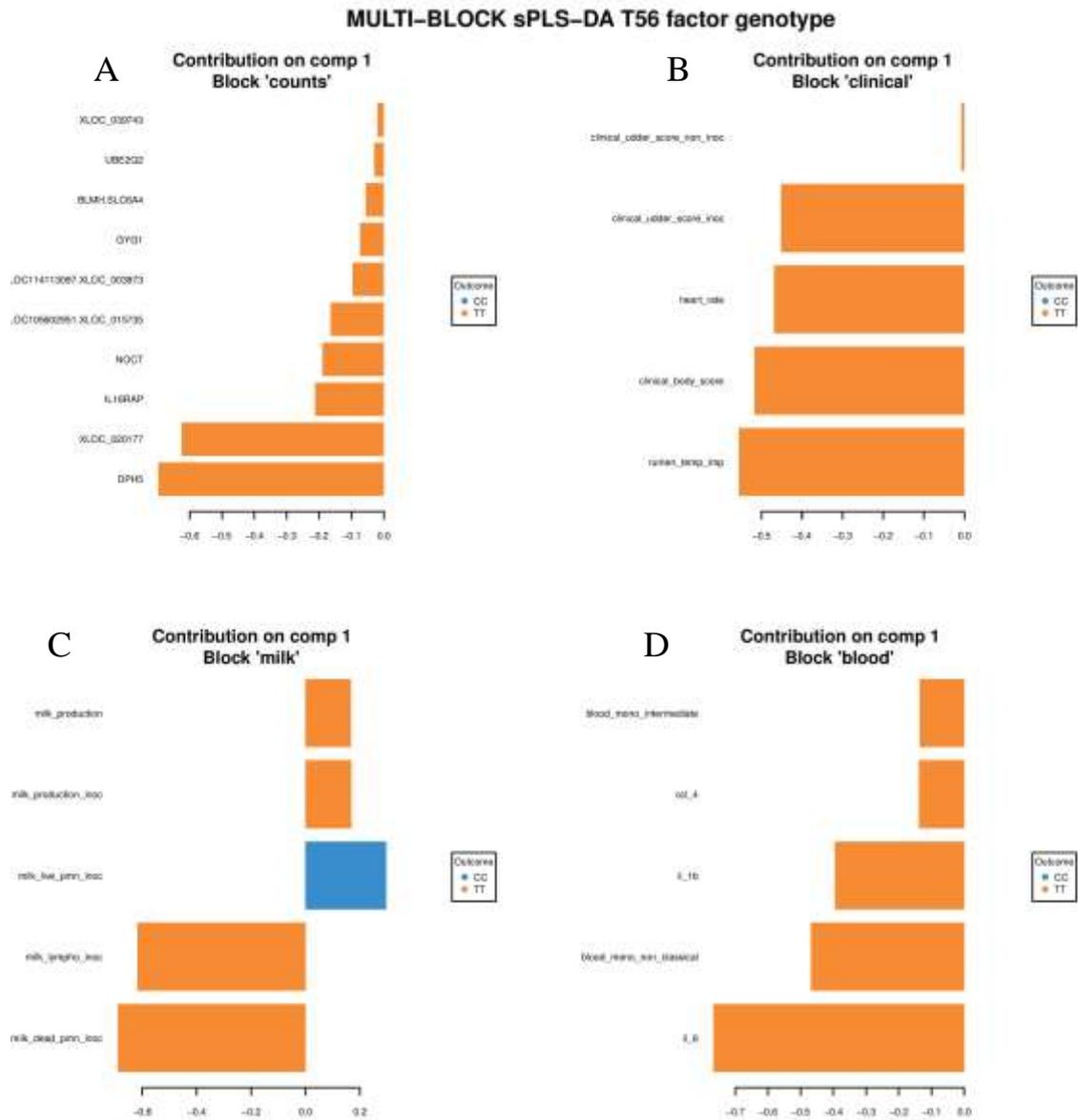
Chaque génotype est matérialisé par une couleur et un symbole différents, comme indiqué dans la légende. Chaque point symbolisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole qui est ici T56.

Annexe 18 Graphique des individus de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype SOCS2 » dans les blocs de données de comptages de gènes (A), de données cliniques (B), de données du lait (C) et de données du sang (D) à T56 de l'infection expérimentale ([Chapitre IV.](#))



Chaque génotype est matérialisé par une couleur différente, comme indiqué dans la légende. Chaque individu labellisé représente un échantillon, correspondant donc à un animal à un temps de contrôle du protocole qui est ici T56.

Annexe 19 Variables expliquant le plus l'axe 1 de la sparse multi-bloc PLS-DA avec comme facteur discriminant le « génotype *SOCS2* » dans les blocs de données de comptages de gènes (A), de données cliniques (B), de données du lait (C) et de données du sang (D) à T56 de l'infection expérimentale ([Chapitre IV.](#))



Alternative methods improve the accuracy of genomic prediction using the information of a pleiotropic point mutation in a dairy sheep model

Claire Oget¹, Marc Teissier¹, Jean-Michel Astruc², Gwenola Tosser-Klopp^{1*} and Rachel Rupp^{1*}

¹ GenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France

² Institut de l'Élevage, 31321, Castanet-Tolosan, France

* These authors contributed equally to this work

claire.oget@inra.fr (Corresponding Author)

Abstract

We evaluated genomic prediction accuracy using different methods to include the effect of a major gene pleiotropic point mutation. We studied as an example the *SOCS2* gene mutation influencing the host's inflammatory response to mammary infections, body growth and milk production.

We performed genetic evaluation for Somatic Cell Scores as well as for four milk production and three mammary type traits. The data was composed of 14,729 genotyped animals (Illumina OvineSNP50 BeadChip). Pedigree-based BLUP was compared with single-step GBLUP (ssGBLUP) and four weighted ssGBLUP methods (WssGBLUP), which give weights to SNPs according to their effect on the evaluated trait. In order to include the *Socs2* genotype information in the evaluation, we previously imputed the point mutation for all the genotyped individuals. We then also tested pedigree and genomic-based Gene Content methods, a multiple-trait genomic model that includes *Socs2* genotypes as a trait, and all the ssGBLUP methods previously described when adding the point mutation as a SNP among markers. We computed accuracies of genomic estimated breeding values (GEBV).

We gained +12.4% in GEBV average accuracy between pedigree-based BLUP and ssGBLUP models and +3.4% between ssGBLUP and WssGBLUP methods. The inclusion of the *Socs2* genotypes with the genomic Gene Content method gave similar results as WssGBLUP. The addition of the *Socs2* SNP in ssGBLUP methods only brought an average gain of +0.23%. Weighted ssGBLUP methods were therefore more efficient for detecting SNPs with strong effects and for better predicting GEBV than ssGBLUP. When no chip data is available for genomic selection, the Gene Content method can efficiently account for partial genotyping information on major genes (+4.35%).

Keywords

Genomics; Genomic Selection; Quantitative Trait Loci; Genome-Wide Association Study; Dairy Sheep; Mastitis; Disease Resistance;

PLEIOTROPIC EFFECTS OF THE R96C MUTATION IN THE DAIRY SHEEP *Socs2* GENE

Claire Oget¹ - Advisors: Rachel Rupp¹, Gwenola Tosser-Klopp¹
¹ INRA, UMR 1388 Génétique, Physiologie et Systèmes d'Elevage, Castanet-Tolosan, France

CONTEXT AND OBJECTIVE

Mastitis is a serious burden for the milk industry (quality, cost). Genetic selection for improved immunity against mammary infections and resistance to mastitis is now implemented in breeding programs of several dairy breeds.
Resistance to mastitis is a highly complex trait which genetically-determined biological basis is mostly unknown.
 Rupp et al. (2015) identified a **causal mutation controlling mammary inflammation** in sheep using a GWAS (Genome Wide Association Study) for the milk somatic cell count (SCC) in dairy sheep, and highlighted the **major role of *Socs2*** as a potential tradeoff factor between host inflammatory response to mammary infections, and milk production.
Objective: Study at the ovine population- and animal-scale of the effect of *Socs2* genotypes on production and health

METHODOLOGY & RESEARCH STRATEGIES

Task 1: Association of the *Socs2* genotype with mastitis and production traits in commercial and experimental sheep populations

- Managing the *Socs2* mutation in the Lacauve population
- GWAS with individuals from INRA UE La Fage

Task 2: Prospective study on the management of the *Socs2* mutation in the context of genomic selection

- Testing methods to include the *Socs2* gene information in genomic evaluation

Task 3: Response to an experimental challenge in sheep carrying the different *Socs2* genotypes

- Experimental infections
- Transcriptome profiling
- Statistical analyses
- Biological analysis and interpretation

Task 4: *In silico* screening of *Socs2* sequences and genes modified within the *Socs2* regulome

- Systematic screening of *Socs2* polymorphisms in other dairy ruminant species

EDUCATION

Cursus Master's Degree in Engineering, Agronomy, INPT ENSAT, Toulouse
Specialization in farming systems and livestock products

Master Subject « Genomic signatures of evolutionary history of French caprine breeds »
Advisors: Isabelle Palière¹, Bertrand Servin¹

Funding of PhD ANR Project REIDSOCs (Robustness, Efficiency, Inflammation and Diseases under the control of *Socs2*)
- 20th PhD seminar of the Genetic Animal division - Rennes, May 9 & 10, 2017

Centre Ecologie-Evolution
Génétique des Populations
13132021 - 13132022 Centre National de la Recherche Scientifique
INRA - UR 1313

INRA - UR 1313
Génétique des Populations
13132021 - 13132022

FINE MAPPING OF 7 QTL REGIONS IN DAIRY SHEEP CONFIRMS PLEIOTROPIC EFFECT OF THE R96C MUTATION IN THE *Socs2* GENE ON SCC, BACTERIAL INFECTION, SIZE AND MILK PRODUCTION



INRA
SCIENCE & IMPACT

C. Oger¹, C. Alain¹, D. Portes², G. Foucras³, G. Tasser-Blapp⁴ & B. Rupp^{1*}

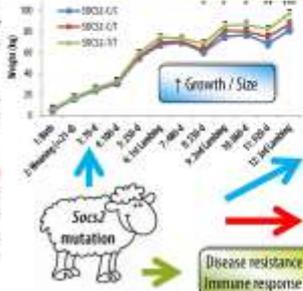
¹ GenPhySE, Université de Toulouse, INRA, INP, Castanet-Tolosan, France; ² INRA, Domaine de La Fage, 12250 Saint-Jean-et-Saint-Paul, France; ³ INRAE, Université de Toulouse, INRA, INP, Castanet-Tolosan, France; *These authors contributed equally to this work.

Discovery of the *Socs2* mutation

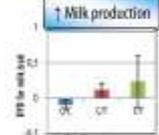
Mastitis is a serious burden for the milk industry (quality, cost). Genetic selection for improved immunity against mammary infections and resistance to mastitis is now implemented in breeding programs of several dairy breeds.

Resistance to mastitis is a highly complex trait which genetically-determined biological basis is mostly unknown.

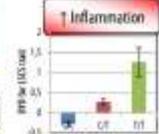
Rupp et al. (2015) identified a causal mutation controlling mammary inflammation in sheep using a GWAS (Genome Wide Association Study) for the milk somatic cell count (SCC) in dairy sheep, and highlighted the major role of *Socs2* as a potential tradeoff factor between host inflammatory response to mammary infections, and milk production.



↑ Growth / Size



↑ Milk production



↑ Inflammation



GWAS using a 960 custom-designed ovine SNP chip confirms strong effect of the *Socs2* mutation

504 Lacaune ewes
INRA experimental facility of La Fage (UE 321, Roquefort, France)

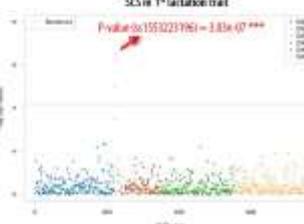
Genotyped with the 35R-mastitis-960-SNP chip, on chromosomes (QARI) 2, 3, 5, 16 and 18

SNP ss1553223196:

- location of the *Socs2* mutation
- QARI3, position 129,722,200 (v3,1)

Phenotyped :
Milk Somatic Cell Score (SCS), milk production, weights, mammary abscess & *Staphylococcus* spp. abundance in milk

Trait	P-value ss1553223196
SCS in 1 st lactation trait	3.03e-07 ***
Weight at day 250	3.71e-04 *
<i>Staphylococcus</i> spp. abundance in milk in 1 st lactation	6.80 e-04 *
Mammary abscess	2.83e-03 *
Milk production in 1 st lactation	2.01e-01 *

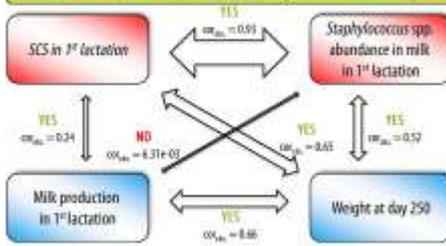


Manhattan plot for SCS in 1st lactation trait

Effect of the *Socs2* gene :

↑ Inflammatory response
↑ Animal size & Intramammary infections
↓ Milk production (no effect here)

Pleiotropic effect of the *Socs2* gene confirmed by the « Close Linkage (vs) Pleiotropism » (CLIP) test



Method: The CLIP test (David et al., 2013) compares two traits and rejects the hypothesis of a pleiotropic QTL if the square of the observed correlation between a combination of apparent effects at the marker level is below the minimal value it can take under the pleiotropic assumption.

Strong correlations between all these Quantitative Trait Loci (QTL), even with milk production
= **pleiotropy of the *Socs2* gene**

Hypothesis of an indirect role of *Socs2* on milk production through its effect on growth

World Congress on Genetics Applied to Livestock Production

Auckland, New Zealand, 11-16 February 2018



INRA
SCIENCE & IMPACT



GenPhySE



ANR



Fine mapping of 7 QTL regions in dairy sheep confirms pleiotropic effect of the R96C mutation in the *Socs2* gene on SCC, bacterial infection, size and milk production

C. Oget¹, C. Allain¹, D. Portes², G. Foucras³, G. Tosser-Klopp^{1*} & R. Rupp^{1*}

¹ GenPhySE, Université de Toulouse, INRA, INPT, ENVT, Castanet Tolosan, France
claire.oget@inra.fr / Riberc@univit.com (Corresponding Author)

² INRA, Domaine de La Fage, 12250 Saint-Jean-et-Saint-Paul, France

³ IHAP, Université de Toulouse, INRA, INPT, ENVT, Castanet Tolosan, France

* These authors contributed equally to this work.

Summary

The aim of this study was to investigate the pleiotropic effects of a *Socs2* gene point mutation associated to lower resistance to mastitis, an inflammation of the mammary gland, mostly due to bacterial infection by Staphylococci in dairy sheep. GWAS were performed on SCC (Somatic Cell Count), bacterial infection, animal size and milk production using a 960 custom-designed ovine SNP chip and a data set of 504 Lacaune dairy ewes. The results confirmed the strong effect of this mutation on the inflammatory response (p-value of LSCS_L1 = 3.03e-07 and p-value of abscess = 2.83e-03), a slighter effect on animal size (p-value of W_DAY_250 = 3.71e-04) and also showed for the first time an effect of this mutation on intramammary infections (p-value of STAPH_L1 = 6.80e-04). In contrast, no effect on milk production could be confirmed (p-value of MILK_L1 = 2.01e-01), suggesting an indirect role of SOCS-2 protein on milk production. Then pleiotropy tests (CLIP test) proved pleiotropic effects of the mutation on SCC, bacterial infection, size and more indirectly on milk production.

Keywords: dairy sheep, genetic resistance, mastitis, genome wide association study

Introduction

Mastitis is an inflammation of the mammary gland, which is mostly due to bacterial infections by Staphylococci in dairy sheep. Mastitis is a serious burden for the milk industry due to the altered quality of milk and increased cost for production and flock renewal. Beside hygienic measures, genetic selection for improved resistance to mastitis is now considered and implemented in breeding programs of several dairy breeds worldwide. However, resistance to mastitis is a highly complex trait which genetically-determined biological basis is mostly unknown.

In dairy sheep, several QTL regions controlling milk somatic cell count (SCC), an indirect indicator of udder infection, have been identified within the EU-funded 3SR project (Sustainable Solutions for Small Ruminants, FP7-KBBE-245140). These QTL regions were then confirmed by Banos *et al.* (2017) with four mastitis indicator traits (clinical mastitis occurrence, milk SCC, total viable bacterial count in milk and the California mastitis test). Amongst these QTLs, Rupp *et al.* (2015) identified a SNP (Single-Nucleotide Polymorphism) in the coding frame of *Socs2* associated with high SCC in the Lacaune breed. This point mutation was localized in the SH2 binding domain of the SOCS-2 protein leading to the loss

of ligand recognition. This strongly suggests that the identified R96C mutation was causal, although the effect of other SNPs in the QTL region could not be fully excluded. It was also demonstrated that size, weight and milk production were significantly increased in ewes carrying the mutated *Socs2* allele when compared to wild type sheep.

The objective of this work was (i) to confirm and fine map seven ovine QTLs controlling mastitis resistance using a 960 custom-designed ovine SNP chip, (ii) to assess the effect of confirmed QTLs on bacterial infection, and (iii) to identify and prove possible pleiotropic effects of QTLs on animal size and milk production.

Material and methods

Animals and phenotypes

The data set included 504 Lacaune dairy ewes from a divergent selection based on extreme breeding values for the somatic cell score (SCS) at the INRA experimental facility of La Fage (UE 321, Roquefort, France) (Rupp *et al.*, 2009).

Monthly SCC were log-transformed for normality into SCS. Test-day SCS and milk yield were averaged over lactation and corrected for the year of control and feeding method. The considered variables were LSCS and MILK for the first and second lactation (L1 and L2, respectively).

The *Staphylococcus* spp. abundance in milk was quantified at three time points in first lactation by a qPCR-based technique developed at IHAP laboratory (ENV Toulouse). Briefly, the method combines an original extraction technique (milk centrifugation in 96-well plates, enzymatic lysis and semi-automated DNA extraction) and high throughput qPCR in 384-well format. Results were expressed as a bacterial titre (quantity of equivalent bacterial genomes in 10 µL of milk), expressed on a logarithmic scale (STAPH_L1). All results from each ewe were averaged and corrected for the effects of month and year of sampling.

Chronic mastitis was also assessed by the presence of mammary abscess, recorded by clinical examination. Animals were recorded as “1” (case) when examinations detected the presence of at least one abscess at least twice, while animals were recorded “0” (control) when they were examined healthy (without any abscess) at least three times.

Each ewe was weighed at birth (W_BIRTH), and at the age of 100 (W_DAY_100) and 250 days (W_DAY_250), and after 1st (on average 412 days, W_1ST_LAMB) and 2nd lambing (on average 744 days, W_2ND_LAMB), and at the age of 920 days (W_DAY_920). The variables of weight were corrected for year and feeding method. Table 1 provides the mean, the standard deviation and headcounts for each of these phenotypes.

Genotyping with a mastitis-specialized 960-SNP chip

All animals were genotyped with a 960 custom-designed ovine SNP chip (http://genoweb.toulouse.inra.fr/~tosser/3SR-WP3-960_snp_mastitis/). The chip was designed and developed within the “3SR” EU project (http://cordis.europa.eu/project/rcn/95054_en.html) based on several QTLs associated with SCC that were found in Spanish Churra, French Lacaune and Italian Sardinian-Lacaune Backcross populations. Seven regions of interest were selected based on commonalities among populations or on high significance on sheep chromosomes (OAR) 2, 3, 5, 16 and 18. SNPs were selected from the 54K or 800K Illumina ovine chips (Nicolazzi *et al.*, 2015) or were identified within the “3SR” project by novel genome sequencing. Details are provided

in Table 2. Genomic positions refer to *Ovis aries* assembly v3.1 (reference NCBI).

Genome wide association mapping

After quality control, 755 SNPs were kept for the study. Genome-wide association studies (GWAS) were performed for each phenotype using the univariate mixed model approach implemented in the Genome-wide Efficient Mixed Model Association (GEMMA) software (Zhou & Stephens, 2012). A Bonferroni correction of $\alpha = 5\%$ was applied (significance threshold = $\alpha/\text{number of SNP}$). SNPs with a p-value $< 6.63\text{e-}05$ were considered to be significantly associated at the chip level. Chromosome-wide significance thresholds were also calculated with the same method.

Close Linkage versus Pleiotropism Test

Because QTL hits were found in the same genomic region for different traits (OAR3 and OAR16), the “Close Linkage versus Pleiotropism test” (CLIP) developed by David *et al.* (2013) was implemented in order to try and distinguish between pleiotropy (a single QTL affecting more than one trait) and/or close linkage (different QTLs that are physically close). Briefly, the test compares two traits and rejects the hypothesis of a pleiotropic QTL if the square of the observed correlation between a combination of apparent effects at the marker level is below the minimal value it can take under the pleiotropic assumption.

Results and Discussion

Genome wide association mapping

The OAR3 region (around 129 Mb) was found significant for 9 out of 12 traits related to SCC, bacterial titre and weight. Table 1 provides for each analysis the p-values of the two most significant SNPs. The first SNP, ss1553223196, is in the coding frame of the *Socs2* gene at position 129,722,200 and the second, ss1553223200, is located in the 3'UTR of CRADD gene at position 129,927,538. The p-values of the *Socs2* SNP were strongly significant for the LSCS traits (p-values equal to $3.09\text{e-}07$ for the 1st lactation), and, to a lesser extent, for weight phenotypes (p-values from $3.71\text{e-}04$ to $1.54\text{e-}03$). These results confirmed earlier findings (Rupp *et al.*, 2015) in a new independent population. The present study further showed a significant association of the OAR3 QTL region with bacterial titre and chronic mastitis. Finally, milk production traits were not found associated to any of the SNPs in this data set.

The SNP located in the CRADD gene was always less significant than the SNP in the *Socs2* gene, except for W_DAY_920. This is additional evidence that the *Socs2* gene is the causal mutation for mastitis resistance and weight, as strongly suggested by Rupp *et al.* (2015).

Some other SNPs from the 960-custom mastitis chip which are not on OAR 3, were found significant in our study: one SNP on OAR 16 for LSCS_L2 (p-value = $9.05\text{e-}07$; MAF = 0.015), another SNP on OAR 19 for W_DAY_920 (p-value = $3.43\text{e-}05$; MAF = 0.143), another SNP on OAR 16 for MILK_L1 (p-value = $1.87\text{e-}05$; MAF = 0.062), and one SNP on OAR 19 for MILK_L2 (p-value = $1.58\text{e-}04$; MAF = 0.219). These 4 SNPs, however, were isolated hits on a single trait and gave low confidence for true QTLs.

Pleiotropic effects of the mutation

The CLIP test was performed for LSCS_L1, STAPH_L1, W_DAY_250 and MILK_L1 traits in the OAR 3 QTL region (13 SNPs ranged from 129.68 Mb to 130.10 Mb) in order to determine whether those QTLs were a single pleiotropic QTL or different QTLs in close linkage. The results are presented in Table 3. The pleiotropic assumption was never rejected in any of the two-trait analyses with LSCS_L1, STAPH_L1 and W_DAY_250, thanks to strong observed correlations of SNP effects, especially between LSCS_L1 and STAPH_L1 traits ($r = 0.93$). This suggests that all these traits may be controlled by the same causal mutation. In contrast, as soon as we included MILK_L1 trait, the observed correlations dropped down except for the analysis with W_DAY_250 and the pleiotropic assumption was even rejected for the analysis with STAPH_L1.

Our results strongly suggest a direct pleiotropic effect of the *Socs2* gene mutation on the inflammatory response, but also on the control of the infection. This can be explained by changes in the regulation of signal transduction from several receptors like the well-described Janus Kinase (JAK)/signal transducers and activators of the transcription (STAT) pathway. The direct effect of the mutation on animal weight is also demonstrated in our study. The interaction between SOCS-2 protein and the growth hormone receptor and signalling may explain this effect. In contrast, the small effect of the *Socs2* gene mutation on milk production found by Rupp *et al.* (2015) is not confirmed in any of the analyses with our data set. The differences observed in milk production between animals carrying or not the *Socs2* mutation could have been caused by an indirect effect on animal size rather than by a direct interaction between SOCS-2 protein and the prolactin hormone receptor.

Conclusion

Our study confirms one ovine QTL controlling mastitis resistance on chromosome 3 using a 960 custom-designed ovine SNP chip. GWAS and CLIP test were performed and confirmed that the QTL is caused by the R96C mutation in the *Socs2* gene with pleiotropic effects on SCC, bacterial infection, size and more indirectly on milk production.

List of References

- Banos G., G. Bramis, S. J. Bush, E. L. Clark, M. E.B. McCulloch, J. Smith, G. Schulze, G. Arsenos, D. A. Hume, A. Psifidi, 2017. The genomic architecture of mastitis resistance in dairy sheep. *BMC genomics*. 18: 624.
- David, I., J. M. Elsen, D. Concordet, 2013. CLIP Test: a new fast, simple and powerful method to distinguish between linked or pleiotropic quantitative trait loci in linkage disequilibria analysis. *Heredity*. 110(3): 232-238.
- Nicolazzi, E. L., A. Caprera, N. Nazzicari, P. Cozzi, F. Strozzi, C. Lawley, A. Pirani, C. Soans, F. Brew, H. Jorjani, G. Evans, B. Simpson, G. Tosser-Klopp, R. Brauning, J. L. Williams, A. Stella, 2015. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genomics*. 16: 283.
- Rupp, R., D. Bergonier, S. Dion, M. C. Hygonenq, M. R. Aurel, C. Robert-Granié, G. Foucras, 2009. Response to somatic cell count-based selection for mastitis resistance in a divergent selection experiment sheep. *J. Dairy Sci*. 92(3): 1203-1219.
- Rupp, R., P. Senin, J. Sarry, C. Allain, C. Tasca, L. Ligat, D. Portes, F. Woloszyn, O. Bouchez, G. Tabouret, M. Lebastard, C. Caubet, G. Foucras, G. Tosser-Klopp, 2015. A point mutation in suppressor of cytokine signalling 2 (*Socs2*) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet*. 11(12): e1005629.
- Zhou, X. & M. Stephens, 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet*. 42: 821-824.

Additional Figures and Tables

Table 1. Mean, Standard Deviation, Headcount and p-values of the two most significant SNPs located on chromosome 3 for each trait.

Trait	Mean (\pm SD)	Nb	P-value ¹	
			ss1553223196	ss1553223200
LSCS_L1	3.08 \pm 1.57	494	3.03e-07***	5.24e-07***
LSCS_L2	3.33 \pm 1.92	391	2.61e-05***	6.41e-05***
ABSCCESS	15 cases / 361 controls	376	2.83e-03*	2.24e-02
STAPH_L1	0.69 \pm 0.83	404	6.80e-04*	1.61e-03*
W_BIRTH	3.9 \pm 0.6	502	3.92e-02	4.62e-02
W_DAY_100	28.7 \pm 3.0	492	5.97e-04*	1.17e-03
W_DAY_250	50.6 \pm 4.1	472	3.71e-04*	2.27e-03
W_1ST_LAMB	64.7 \pm 6.7	438	5.01e-04*	1.83e-03
W_2ND_LAMB	73.4 \pm 7.7	393	1.54e-03*	2.60e-03
W_DAY_920	67.3 \pm 7.2	378	9.48e-04*	9.34e-04
MILK_L1	999 \pm 182	494	2.01e-01	2.44e-01
MILK_L2	1 114 \pm 231	391	5.77e-01	7.65e-01

¹***Genome wide: SNP with a p-value < 6.63e-05; *Chromosome wide: 3.85e-03 for chromosome 3

Table 2. SNP content for the 3SR-mastitis-960-SNP chip

OAR	Discovery population(s)	QTL loc (Mb)	Covered interval (Mb)		Origin of the SNPs			SNP number	
			min	max	54K	800K	Seq	before	after
2	Lacaune	125	124.6	125.5	19	10		29	24
2	Sardinian BC	206.7	204.2	209.7	106	146		252	196
	Churra	208.7							
3	Lacaune	129.7	129.7	130.1	4	6	10	20	13
5	Lacaune	96.5	95.8	96.2	11	9		20	17
12	Churra	18.3	17.9	20.4	54	76		130	101
	Sardinian BC	18.4							
16	Lacaune	36.2	28.3	36.8	144	110		254	202
19	Churra	26.3	23.6	28.9	75	180		255	202
	Lacaune	28.6							

Table 3. Results of the CLIP test

Trait 1	Trait 2	Correlation observed	Correlation limit at 5%	Significance (corr obs > corr limit)
LSCS_L1	STAPH_L1	0.93	0.67	Yes
LSCS_L1	W_DAY_250	0.65	0.58	Yes
LSCS_L1	MILK_L1	0.24	0.14	Yes
STAPH_L1	W_DAY_250	0.52	0.19	Yes
STAPH_L1	MILK_L1	6.31e-03	3.58e-02	No
W_DAY_250	MILK_L1	0.66	1.09e-02	Yes

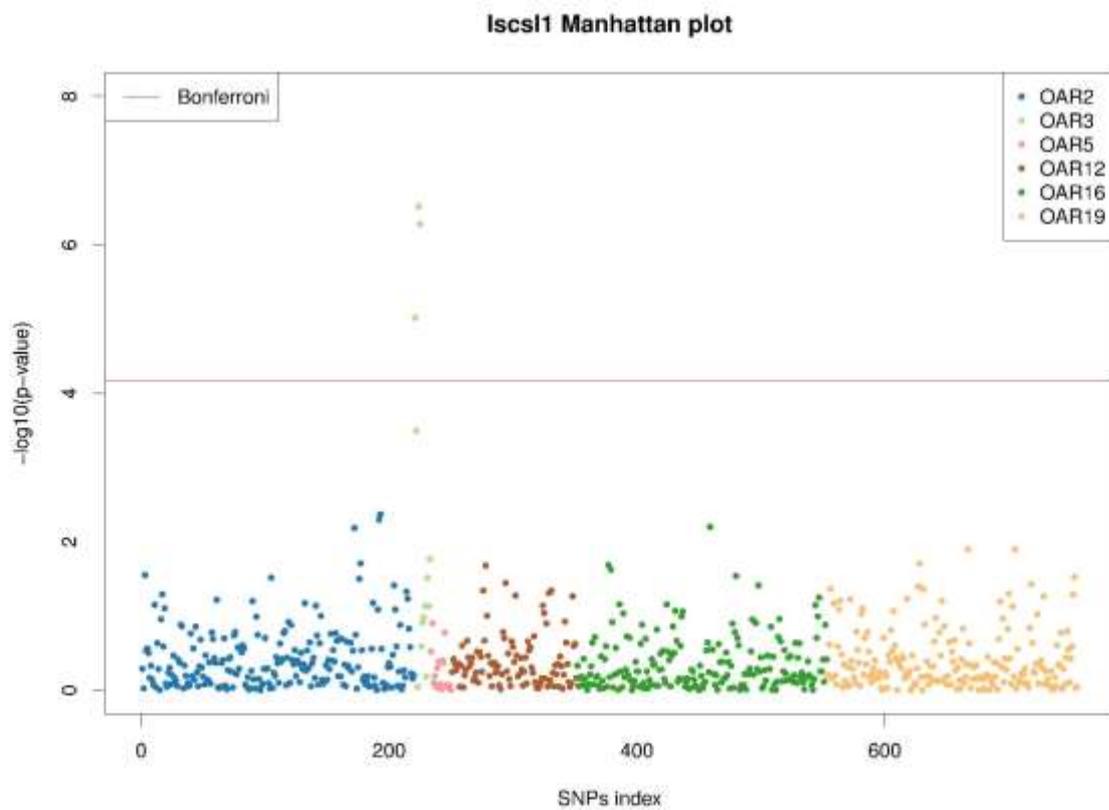


Figure 1. Manhattan plot of the Genome-Wide Association Study on LSCSL1. Each colour change corresponds to a chromosome.

Résumé grand public

Nous avons étudié une mutation naturellement présente dans le gène *SOCS2* (*Suppressor Of Cytokine Signaling 2*) chez la brebis laitière. Cette mutation a un effet pléiotrope : défavorable sur les mammites (infection de la mamelle) et favorable pour les caractères de croissance et de production laitière. Dans cette thèse, nous avons montré que la prise en compte de cette mutation dans les modèles d'évaluation génétique permettait d'envisager une sélection plus efficace pour améliorer la santé mammaire tout en préservant le potentiel de production. Nous avons également mis en évidence de nouvelles associations de cette mutation avec la santé (statut infectieux des mamelles) et la reproduction. Enfin, l'étude de la réponse à un *Staphylocoque* chez des brebis porteuses et non porteuses de la mutation a montré une différence de régulation du système immunitaire en lien avec la protéine SOCS-2, *via* la voie de signalisation JAK/STAT (interférons et STAT3).

General public abstract

We investigated a naturally occurring mutation in the *SOCS2* gene (Suppressor Of Cytokine Signaling 2) in sheep. This mutation has a pleiotropic effect: unfavorable for mastitis (udder infection) and favourable for growth and milk production traits. In this PhD thesis, we showed that including this mutation in genetic evaluation models allowed considering a more effective selection to improve mammary health while preserving the production potential. We have also identified new associations of this mutation with health (udder infection status) and reproduction. Finally, the study of the response to *Staphylococcus* in ewes without and with the mutation showed a difference in the immune system regulation related to the SOCS-2 protein, through the JAK/STAT signalling pathway (interferons and STAT3).

Résumés

Dans cette thèse, nous nous sommes intéressée à une mutation dans le gène *SOCS2* (*Suppressor Of Cytokine Signaling 2*) dont l'effet est pléiotrope, c'est-à-dire qu'il concerne plusieurs caractères d'intérêt chez la brebis laitière Lacaune. Cette mutation est défavorable pour la résistance aux mammites (inflammation de la mamelle), et favorable pour les caractères de croissance et de production laitière. Située dans le domaine de liaison très conservé de la protéine SOCS-2, cette mutation ponctuelle (SNP : *Single Nucleotide Polymorphism*) entraîne la perte de fonctionnalité de la protéine qui est impliquée dans une voie majeure de signalisation chez les mammifères, la voie JAK (*Janus Kinase*) / STAT (*Signal Transducer and Activator of Transcription*), régulant un large spectre de cytokines et de facteurs de croissance.

Le premier objectif de cette thèse était de proposer la prise en compte de cette mutation dans le contexte de la sélection génomique dans la race Lacaune. Les méthodes d'évaluation pondérées, et l'ajout du SNP *SOCS2* parmi les marqueurs de la puce, ont apporté des gains de précisions sur les prédictions (respectivement +3,99% et +0,26% en moyenne), laissant envisager la possibilité d'une sélection plus efficace. La méthode *Gene Content* était également intéressante car elle a permis de dissocier la valeur génétique due au gène *SOCS2* de celle provenant des autres gènes (polygène). Ainsi nous avons montré que la sélection actuelle en race Lacaune permet de diminuer la fréquence de l'allèle défavorable *SOCS2* tout en améliorant la résistance aux mammites expliquée par le reste du génome.

Le deuxième objectif était d'étudier les effets de SOCS-2 sur des caractères d'intérêt non évalués jusqu'à présent : statut infectieux des mamelles et reproduction chez la femelle, et croissance des mâles. Des études d'association avec une puce de 960 SNP, dont le SNP *SOCS2*, ont confirmé l'effet de *SOCS2* sur la production laitière et sur l'inflammation mammaire, et ont montré un effet direct sur l'infection à l'aide de nouveaux phénotypes fins (bactériologie, examens cliniques). Des modèles linéaires ont montré que la mutation *SOCS2* était associée à une augmentation de poids chez les jeunes mâles (+1,5%), comme cela avait été observé chez la brebis, et à une diminution du taux de réussite à l'insémination artificielle chez les brebis (+1,3 fois de risque d'échec), sans effet sur la prolificité. Ces résultats marquants et originaux sur la reproduction ouvrent la voie à des analyses complémentaires pour déterminer notamment à quel stade biologique (ovulation, fécondation, implantation embryonnaire, etc.) la protéine SOCS-2 pourrait intervenir.

Le dernier objectif était de déterminer le rôle de la protéine SOCS-2 dans les mécanismes biologiques sous-jacents de l'immunité grâce à l'étude de la réponse de 14 brebis porteuses et non porteuses de la mutation à l'état homozygote dans le cadre d'une infection intramammaire par un *Staphylococcus aureus*. L'ensemble des brebis a présenté deux pics de recrutement de cellules immunitaires vers la mamelle à 20 et 88 heures post-inoculation (T20 et T88). Une analyse différentielle des gènes exprimés a confirmé une modification du transcriptome à partir de T16 en lien avec l'activation du système immunitaire, des voies de réparation de l'ADN et d'apoptose cellulaire. Les brebis portant la mutation *SOCS2* se sont différenciées des brebis sauvages, par une réaction clinique accrue à T56 et une différence de recrutement de lymphocytes T au moment des pics inflammatoires. A T56, un total de 177 gènes était différenciellement exprimé entre les deux génotypes. Une analyse fonctionnelle de ces gènes, et l'intégration entre l'expression génique et les phénotypes, suggèrent que la mutation conduit à une sur activation des voies de transmission de signal et de régulation (interférons, STAT3) avec un impact délétère sur l'état clinique des brebis.

In this PhD thesis, we investigated a mutation in the *SOCS2* gene (*Suppressor Of Cytokine Signaling 2*) with a pleiotropic effect, *i.e.* it affects several traits of interest in the Lacaune dairy sheep. This mutation is unfavorable for mastitis resistance (udder inflammation), and favorable for growth and milk production traits. This point mutation (SNP - *Single Nucleotide Polymorphism*), located in the highly conserved binding domain of the SOCS-2 protein, causes the loss of functionality of the protein that is involved in a major signaling pathway in mammals: the JAK (*Janus Kinase*) / STAT (*Signal Transducer and Activator of Transcription*) pathway, regulating a large spectrum of cytokines and growth factors.

The first objective of this PhD thesis was to provide a basis for considering this mutation in the context of genomic selection in the Lacaune breed. Weighted evaluation methods, and the addition of the *SOCS2* SNP among the chip markers, brought gains in accuracy on predictions (+3.99% and +0.26% respectively on average), suggesting the possibility of more efficient selection. The *Gene Content* method was also interesting because it allowed dissociating the genetic value due to the *SOCS2* gene from that of the other genes (polygene). Thus, we have shown that the current selection in Lacaune breed allows reducing the frequency of the unfavorable *SOCS2* allele while improving the resistance to mastitis explained by the remaining part of the genome.

The second objective was to study the effects of *SOCS2* on traits of interest not yet investigated: udder infection status and reproduction in females, and growth in males. Association studies with a 960 SNP chip, including *SOCS2* SNP, confirmed the effect of *SOCS2* on milk production and mammary inflammation, and showed a direct effect on infection using new fine phenotypes (bacteriology, clinical examinations). Linear models showed that the *SOCS2* mutation was associated with an increase in weight in young males (+1.5%), as observed in ewes, and a decrease in the artificial insemination success rate in ewes (+1.3 times the risk of failure), with no effect on prolificacy. These striking and original results on reproduction pave the way for further analyses to determine in particular at what biological stage (ovulation, fecundation, embryonic implantation, etc.) the SOCS-2 protein could be involved.

The last objective was to determine the role of the SOCS-2 protein in the underlying biological mechanisms of immunity by investigating the response of 14 homozygous carrier and non-carrier ewes to intramammary infection by *Staphylococcus aureus*. All ewes developed two peaks of immune cell recruitment to the udder at 20 and 88 hours post-inoculation (T20 and T88). Differential analysis of the expressed genes confirmed a modification of the transcriptome from T16 related to the activation of the immune system, the DNA repair and the cellular apoptosis pathways. Ewes carrying the *SOCS2* mutation were distinguished from wild ewes by an increased clinical response at T56 and a difference in T lymphocyte recruitment at the time of the inflammatory peaks. At T56, a total of 177 genes were differentially expressed between the two genotypes. A functional analysis of these genes, and the integration between gene expression and phenotypes, suggest that the mutation leads to an over activation of signal transmission and regulatory pathways (interferons, STAT3) with a deleterious impact on the clinical status of ewes.